

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 08:15:45 ; Search time 23.19 Seconds
(without alignments)
887.985 Million cell updates/sec

Title: US-09-866-034-2
Perfect score: 1399
Sequence: 1 MHKAGLLGLCARAMNSVRMA.....DASYINGEYVVGCTPSRL 278

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1399	100.0	278	21	PRO1800, a Hep27 h
2	1395	99.7	278	22	AAV96729
3	1327	94.9	278	21	AAV693414
4	649.5	46.4	183	21	AAV68735
5	588.5	42.1	276	21	AAV58822
6	588.5	42.1	276	21	AAV58822
7	428	30.6	103	22	AAV54422
8	404	28.9	248	21	AAV54422
9	370	26.4	79	21	AAV54422
10	353	25.2	262	21	AAV54422
11	349	24.9	292	22	AAV54422

12	349	24.9	295	22	AAV9400	Corynebacterium g1
13	349	24.9	295	22	AAV9401	Corynebacterium g1
14	348	24.9	246	21	AAV15707	Staphylococcus aur
15	348	24.9	246	22	AAV02195	S. aureus NADPH-de
16	345	24.7	253	21	AAV31819	Arabidopsis thalia
17	341	24.4	243	22	AAV01032	Arabidopsis thalia
18	340	24.3	243	19	AAV80670	CPE 35 protein seq
19	340	24.3	243	19	AAV80670	S. pneumoniae seq
20	339	24.2	254	22	AAV49773	Streptococcus pneu
21	339	24.2	254	22	AAV49773	protein with acety
22	336	24.0	306	22	AAV81644	S. epidermidis ope
23	333.5	23.8	251	20	AAV36959	Chlamydia trachoma
24	333.5	23.8	251	20	AAV36959	S. epidermidis ope
25	331.5	23.7	261	21	AAV83032	Chlamydia trachoma
26	331.5	23.7	261	21	AAV83032	S. epidermidis ope
27	331.5	23.7	261	21	AAV83032	S. epidermidis ope
28	330	23.6	271	21	AAV32504	B. subtilis glucos
29	330	23.6	271	21	AAV32504	Amino acid sequenc
30	328.5	23.5	315	17	AAV43270	Arabidopsis thalia
31	328.5	23.5	315	17	AAV43270	Arabidopsis thalia
32	327.5	23.4	340	21	AAV10740	Arabidopsis thalia
33	327.5	23.4	340	21	AAV10741	Arabidopsis thalia
34	326.5	23.3	261	13	AAV27757	Rape leaf beta-ket
35	326.5	23.3	261	13	AAV27757	B. megaterium gluc
36	325.5	23.3	279	21	AAV49315	H. ghiliani/B. me
37	325.5	23.3	261	13	AAV80590	Glucose dehydrogen
38	324	23.2	262	13	AAV27756	Arabidopsis thalia
39	324	23.2	262	13	AAV27756	Sequence of glucos
40	324	23.2	262	13	AAV27756	NAD affinity glucos
41	323.5	23.1	271	21	AAV86337	lytic dehydrogen
42	323.5	23.1	271	21	AAV86337	G. suboxydans DSM
43	323.5	23.1	264	21	AAV32505	Arabidopsis thalia
44	323.5	23.1	264	21	AAV32505	Arabidopsis thalia
45	323.5	23.1	318	21	AAV29312	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AAV96729	standard; Protein: 278 AA.
AC	AAV96729;	
XX	26-SEP-2000	(first entry)
DE	PRO1800, a Hep27 homologue.	
XX	PRO1800; Hep27; homologue; short-chain alcohol dehydrogenase; SCAD;	
KW	secreted protein; transmembrane protein; recombinant production;	
KM	gene therapy.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Peptide	1..15
FT	Modified-site	/label= Signal_peptide
FT	Modified-site	43..48
FT	Modified-site	/note= "N-myristoylation site"
FT	Modified-site	80..85
FT	Modified-site	/note= "N-myristoylation site"
FT	Modified-site	183..186
FT	Modified-site	/note= "N-glycosylation site"
FT	Modified-site	191..196
FT	Modified-site	/note= "N-myristoylation site"
FT	Modified-site	213..218
FT	Modified-site	/note= "N-myristoylation site"
FT	Modified-site	272..277
FT	Modified-site	/note= "N-myristoylation site"
FT	Modified-site	276..278
FT	Peptide	/label= Targeting_signal
FT		/note= "Microbody C-terminal targeting signal"

XX WO200036102-A2.
 XX 22-JUN-2000.
 XX 01-DEC-1999; 99WO-US28634.
 XX 16-DEC-1998; 98US-0112851.
 XX 16-DEC-1998; 98US-0113145.
 XX 22-DEC-1998; 98US-0113511.
 XX 12-JAN-1999; 99US-0115558.
 XX 12-JAN-1999; 99US-0115565.
 XX 12-JAN-1999; 99US-0115733.
 XX 09-FEB-1999; 99US-0119341.
 XX 10-FEB-1999; 99US-0119537.
 XX 12-FEB-1999; 99US-0119965.
 XX 02-JUN-1999; 99WO-US12252.

XX (GETH) GENENTECH INC;

XX Botstein D, Desnovers L, Ferrara N, Fong S, Gao W, Goddard A;
 PI Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
 PI Wood W;

XX WPI: 2000-433586/37.
 DR N-PSDB; AANA51259.

XX Isolated nucleic acid molecule encodes a PRO polypeptide which is a
 PT transmembrane polypeptide

PS Claim 12; Fig 2; 154pp; English.

XX This is PRO1800, a putative human Hep27 protein homologue, which has
 CC recently been shown to have homology to known short-chain alcohol
 CC dehydrogenase (SCAD) family of proteins. The invention concerns novel
 CC secreted and transmembrane proteins, designated PRO polypeptides. The
 CC cDNA and gene sequences are useful in the recombinant production of PRO
 CC polypeptides, as a hybridization probe to screen libraries to isolate
 CC cDNAs with sequence identity to PRO polypeptides or to map the gene
 CC encoding the PRO polypeptides and analyzing genetic disorders. The
 CC cDNA/gene can also be used to produce transgenic animals useful for the
 CC development and screening of therapeutically useful reagents. They can
 CC also be used in gene therapy, e.g. to replace a defective gene.

XX Sequence 278 AA;

Query Match 100.0%; Score 1399; DB 21; Length 278;
 Best Local Similarity 100.0%; Pred. No. 4,9e-132;
 Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHKAGLLGLCARAWNSYRMASSGWTTRDPLANKVALVTASDGIQFPAIRRLAODGAHV 60
 Db 1 mhkagllglcarawnsyrmasssgwttrdplankvalvtascdgiqfatarrlagdgahv 60
 QY 61 VSSKQONDOAVATLOGEGLSVTVCHWGKAEDREBLVATAVAKLHGIDILVSNAVN 120
 Db 61 vsskqgnvdavatlqgeglsvtgvchvgaedrerlvalavklngidilvsnaavn 120
 QY 121 PFFCSINDVTEWMDKLDINVKAPALMTKAVNPEMKRGGSVYVSSIAESPSGFS 180
 Db 121 pffgsindvteeewdkldlnvkapalmtkavnpemkrggsvlyvsslaafspgfs 180
 QY 181 PYNVSKTALLGLTTLAIELAPRIRVNCIAPGLIKTSFSRMLMDKEESKKEETLRIR 240
 Db 181 pynvsktallgltkltaielaprnirvnciapglktsfstrmlwmdkeeskmkeltlir 240
 QY 241 RLSEPECCATVFLCEDASVYTGTVVVGGSPTSR 278
 Db 241 rlsepedcagivsfllcsedesyiltgetvvvggtpsr 278

RESULT 2

AAB93414
 ID AAB93414 standard; Protein; 278 AA.

XX AAB93414;

XX 26-JUN-2001 (first entry).

DE Human protein sequence SEQ. ID NO:12620.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;

XX WPI: 2001-316749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 12620; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 278 AA;

Query Match 99.7%; Score 1395; DB 22; Length 278;
 Best Local Similarity 99.6%; Pred. No. 1.2e-131;
 Matches 277; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHKAGLLGLCARAWNSYRMASSGWTTRDPLANKVALVTASDGIQFPAIRRLAODGAHV 60
 Db 1 mhkagllglcarawnsyrmasssgwttrdplankvalvtascdgiqfatarrlagdgahv 60

QY	61	VSSRRKOONDDAVNLTQEGSLVSTCTVGHNGKAEDERELVTATVAKLHGCDILVLSNAVN	120
Db	61	VSIRYQNVDDQVAILIGGELISVLTCTVGHVKAEDRELIVATVAKLHGCDILVLSNAVN <th>120</th>	120
QY	121	PEFGSIMVTEVEMWTKTLDINWKAAPALMTKAVVPEMKRGGSVYVTSIAFSPSPGS <th>180</th>	180
Db	121	PEFGSIMVTEVEMWTKTLDINWKAAPALMTKAVVPEMKRGGSVYVTSIAFSPSPGS <th>180</th>	180
QY	181	PYNSKTKALLGLTKTLATIELAPRINRVNCLAPGLIKTFSFSMLWMDKEESMKETLRIR <th>240</th>	240
Db	181	PYNSKTKALLGLTKTLATIELAPRINRVNCLAPGLIKTFSFSMLWMDKEESMKETLRIR <th>240</th>	240
QY	241	RIGEPEDCAGIVSPFCSBDAZYINGETVYVGGGTPSRLL <th>278</th>	278
Db	241	RIGEPEDCAGIVSPFCSBDAZYINGETVYVGGGTPSRLL <th>278</th>	278
RESULT	3		
AY68735			
ID	AY68735	standard; Protein: 278 AA.	
XX	XX	AY68735:	
XX	XX	05-MAY-2000 (first entry)	
De	Short chain alcohol dehydrogenase-related molecule SCRM-1 protein.		
KW	Human: short chain alcohol dehydrogenase-related molecule;		
KW	SCAD-related molecule; SCRM-1; SCRM-2; metabolic regulator;		
KW	cell proliferation regulator; inflammation regulator;		
KW	cell proliferative disorder; immune disorder; arteriosclerosis;		
KW	atherosclerosis; bursitis; cirrhosis; hepatitis; AIDS;		
KW	Addison's disease; adult respiratory distress syndrome; allergy;		
KW	ankylosing spondylitis; amyloidosis; cancer.		
XX			
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
FT	Modified-site	16	
FT	Modified-site	/note- "potential protein kinase C phosphorylation site"	
FT	Modified-site	21	
FT	Modified-site	/note- "potential protein kinase C phosphorylation site"	
FT	Modified-site	25	
FT	Modified-site	/note- "potential casein kinase II phosphorylation site and potential protein kinase C phosphorylation site"	
FT	Modified-site	31	
FT	Modified-site	/note- "potential protein kinase C phosphorylation site"	
FT	Modified-site	33..45	
FT	Modified-site	/note- "SCAD motif"	
FT	Modified-site	34..51	
FT	Modified-site	/note- "glucose/ribitol dehydrogenase motif"	
FT	Modified-site	39..46	
FT	Modified-site	/note- "AMP-binding domain"	
FT	Modified-site	62	
FT	Modified-site	/note- "potential protein kinase C phosphorylation site"	
FT	Modified-site	63	
FT	Modified-site	/note- "potential protein kinase C phosphorylation site"	
FT	Modified-site	108..119	
FT	Modified-site	/note- "glucose/ribitol dehydrogenase motif"	
FT	Modified-site	108..118	
FT	Modified-site	/note- "SCAD motif"	
FT	Modified-site	108..119	
FT	Modified-site	/note- "SCAD motif"	
FT	Modified-site	125	
FT	Modified-site	/note- "potential casein kinase II phosphorylation site"	
FT	Modified-site	156..172	
FT	Modified-site	/note- "glucose/ribitol dehydrogenase motif"	
FT	Modified-site	162..199	
FT	Modified-site	/note- "SCAD motif"	
FT	Modified-site	162..213	
FT	Modified-site	/note- "SCAD signature sequence"	
FT	Modified-site	169..197	

FT	Region	/note="SCAD family signature sequence"
FT	FT	182..201
FT	Domain	/note="glucose/ribitol dehydrogenase motif"
FT	FT	182..186
FT	Region	/note="canonical catalytic site of SCADS"
FT	FT	203..220
FT	Region	/note="glucose/ribitol dehydrogenase motif"
FT	FT	204..213
FT	Modified-site	/note="SCAD motif"
FT	FT	232
FT	FT	/note="potential casein kinase II phosphorylation site and potential protein kinase C phosphorylation site"
FT	Modified-site	236
FT	FT	/note="potential protein kinase C phosphorylation site"
FT	Region	238..258
FT	FT	/note="glucose/ribitol dehydrogenase motif"
PX	WO200004135-A2.	
PX	27-JAN-2000.	
PX	16-JUL-1999:	99MO-US16164.
PX	16-JUL-1998:	98US-O116750.
PX	16-JUL-1998:	98US-O160074.
PA	(INCY-) INCYTE PHARM INC.	
XX	Bandman O, Tang YT, Corley NC, Azimzai Y, Baughn MR;	
XX	WPI; 2000-171266/15.	
DR	N-PADB; AAZ46090.	
PT	New short chain alcohol dehydrogenase polypeptides useful for	
PT	diagnosis, treatment and prevention of cell proliferative disorders	
PT	such as atherosclerosis, cirrhosis and cancers of various tissues -	
PS	Claim 1; Fig 1A-D; 78pp; English.	
CC	The present sequence represents a human short chain alcohol dehydrogenase	
CC	(SCAD)-related molecule designated SCR-1. The specification also	
CC	describes SCR-2. SCR proteins are metabolic, cell proliferation and	
CC	inflammation regulators. The SCR polynucleotides and polypeptides are	
CC	used for treating or preventing a cell proliferative or immune disorder	
CC	in humans. Cell proliferative disorders include arteriosclerosis,	
CC	atherosclerosis, bursitis, cirrhosis, and hepatitis. Immune disorders	
CC	include AIDS, Addison's disease, adult respiratory distress syndrome,	
CC	allergies, ankylosing spondylitis, and amyloidosis. The vectors.	
CC	agonists, antagonists, antibodies and complementary sequences are also	
CC	used for treating the above conditions. The polynucleotides and	
CC	polypeptides are also used for treating cancers of various tissues	
CC	such as adrenal gland, bladder, bone, bone marrow, and brain.	
SO	Sequence 278 AA:	
OY	Query Match	94.9% Score 1327; DB 21; Length 278;
Dd	Best Local Similarity	96.4%; Pred. No. 8, 2e-125;
OY	Matches 268; Conservative	0; Mismatches 10; Indels 0; Gaps 0
OY	1 MHKACILGLCARANMSVSRMSSGMRPRDLAKKVALVTASTGIGFAIRRLAQDGHV	60
Dd	1 mhmarlllglcavarksvrmasmrirtrpltklvalvtastdgifatairrlaqdghvv	60
OY	61 VSSRQGVNDQAVATLGGSLSTGTCTCHVGKAEDRERLVAFAVKLHGIDILVSNAAYN	120
Dd	61 vssrtqgvndqavalatggslstgtctchvkgaedrerlvafavklnhgidilvsnaayn	120
OY	121 PFFGGINVTEEVMPTKDINKAPALMKRAVPPEKEKGSGSVIVSSIAFSPPGFS	180
Dd	121 pfifgindvteevmptktldinkapalmkravppekekgsgsvivssiafsspfgfs	180

Claim 1; Fig 1A-D; 78pp; English.

line present-sequence represents a human short chain alcohol dehydrogenase (SCAD)-related molecule designated SCGM-1. The specification also describes SCGM-2. SCGM proteins are metabolic, cell proliferation and inflammation regulators. The SCGM polynucleotides and polypeptides are used for treating or preventing a cell proliferative or immune disorder in humans. Cell proliferative disorders include artherosclerosis, atherosclerosis, bursitis, cirrhosis, and hepatitis. Immune disorders include AIDS, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, and amyloidosis. The vectors, agonists, antagonists, antibodies and complementary sequences are also used for treating the above conditions. The polynucleotides and polypeptides are also used for treating cancers of various tissues such as adrenal gland, bladder, bone, bone marrow, and brain.

SQ Sequence 278 AA;

Query Match	94.9%	Score 1327	DB 21	Length 278
Best Local Similarity	96.4%	Prod No 8	2010	

Matches	268;	Conservative	0;	Mismatches	10;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	-----	--------	----	------	----

Oy 1 MHKAGLLCLCARANNNSVMASSGMYTRDPLANKVALYASPDIGFAIRRLAODGAHV 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dd 1 mhmarlllgicawarksvmassmrlrtpcltnkvalylvastdgfgafarlrldgrzhv 60

61 VSSRKOQNDQAVATLQEGGLSVTGTVCVHGKAEDRERLVATAVKLHGIDILVSNAAVN 120

Db 61 vssrkqnvdaqavatlqgeglsvtgtvchvgkaedrervatavkllhqiddilvsnaavp 120

QY 121 PFGSIMDVTEEVWDKTLIDINKAPALMTKAVVPMEKRGGSVTVSSIAEFSBCEC 190

Db
121 pfqsimdvtceewdkrtjdinvkanalmtkaywpmcknagawjuzfzcfv

QY 181 PYNVSKTALLGLTTLAIELAPRIRVNCIAPGLIKTSFSRMLMDKEKESMKETLRIR 240
 DB 181 pynvsktallglnltlailaprlrvnciapgliktsfsmldmkeesmketlrlr 240
 QY 241 RLGEPEOCAGIVSFLCEDASYIGETVWVGCGTPSRRL 278
 DB 241 rlgepedcagivsflicsedasyigetvwwvggtpsrll 278

RESULT 4
 ID AAB58822 standard; Protein; 183 AA.
 AC AAB58822;
 DT 27-MAR-2001 (first entry)
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 530.

Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 neurotropic; neuroprotective; antiviral; antiinflammatory; hepatotropic;
 antiinflammatory; antifungal; antiparasitic; cardiant; immune disorder;
 Addison's disease; allergy; autoimmune haemolytic anaemia;
 autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 cardiovascular disorder; wound healing; neurological disease.

XX- Homo sapiens.
 OS
 XX
 PN MO200055173-A1.
 XX
 PL- 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05881.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 DR MPI: 2000-611515/58.
 DR N-PSDB; AAF211725.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 11; Page 963-964; 1299pp; English.

CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neurotropic; neuroprotective; antiviral; antiinflammatory; hepatotropic;
 CC antibacterial; antiinflammatory; antiviral; antiparasitic; cardiant; immune
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis of
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.

XX
 SQ Sequence 183 AA;

Query Match 46.4%; Score 649.5; DB 21; Length 183;
 Best Local Similarity 55.2%; Pred. No. 4e-57;
 Matches 148; Conservative 2; Mismatches 21; Indels 97; Gaps 3;

QY 16 SVRMASGMRDRDPLANKVALVTASTDIGTAIRRLAQDGAHVVSRRKQONQAV-- 73
 DB 8 svrmassmrtrdplnkvavltastldigf-----aspvwprrtprrgqdeaeecg 61
 QY 74 ---ATLOGESLVGTVCVHCVKAEDRELVATVAKLHGCDILVSNAAVNFESINDYT 130
 DB 62 pgggtlqgegslvgtcxxxkaedrelyatavklingidilvsnaavpflgslndvt 121
 QY 131 EEVWDKTLIDINVKAPALMTKAVPEMEKGGSVIVSSIAFSPSPGFSPYVNSKTALL 190
 DB 122 eevwdk----- 127

QY 191 GLTTLAIELAPRIRVNCIAPGLIKTSFSRMLMDKEKESMKETLRIRLGEPEOCAG 250
 DB 128 -----lmdkekesmketlrlrlgepedcag 155

QY 251 IVSFLCEDASYIGETVWVGCGTPSRRL 278
 DB 156 ivsflicsedasyigetvwwvggtpsrll 183

RESULT 5
 ID AAG41219 standard; Protein; 254 AA.
 AC AAG41219;
 DT 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 51256.

XX
 XX
 XX Protein identification: signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 XX
 OS Arabidopsis thaliana.
 OS
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.

XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135121.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137724.
 PR 07-JUN-1999; 99US-0138094.
 PR 08-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139753.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 28-JUN-1999; 99US-0140695.
 PR 29-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0140991.
 PR 01-JUL-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142927.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 06-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 09-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 11-AUG-1999; 99US-0148171.
 PR 12-AUG-1999; 99US-0148319.
 PR 13-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 16-AUG-1999; 99US-0148684.
 PR 17-AUG-1999; 99US-0149368.
 PR 18-AUG-1999; 99US-0149175.
 PR 20-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 23-AUG-1999; 99US-0149723.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149902.
 PR 25-AUG-1999; 99US-0149930.
 PR 26-AUG-1999; 99US-0150566.
 PR 27-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151080.
 PR 31-AUG-1999; 99US-0151303.
 PR 01-SEP-1999; 99US-0151438.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.

[illegible]

PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161404
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161406
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161360
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161920
PR	28-OCT-1999;	99US-0161922
PR	28-OCT-1999;	99US-0161993
PR	29-OCT-1999;	99US-0162142

Query Match 42.1% Score 588.5; DB 21; Length 276;
Best Local Similarity 50.6% Pred. No. 9.5e+51;
Matches 126; Conservative 35; Mismatches 85; Indels 3; Gaps

OY	30 LANKALVLTASTDGGFATARRLADOGAHVVYSRRKQONDOAVTLLOGEGLSYTGTVCH	89
DB	31 Iegkvaivlcastqgijfgtferflegasvsvsrkgaavdeavaklskgldaygilvch	90
OY	90 VGKAEDRELRLVTFAVKLMGCGIDILVSNAVNPFEGSIMDYTEVWDKTLDIRKAPALMT	149
DB	91 vsnaqntrlnlvckltvykygklidivcnaanaapstclpllskskaevldkliweinvksllll	150
OY	150 KAVNVEMERKGSGSVIVISSIAAFSPSPGFSPYNSTKALLGLTKTALIELAPNRIVNC	205
DB	151 qdmaphlek--gsyvlfltslagfpvgamamytkcallgiklaaaemap-dlrva	207
OY	210 LAPGLIKTSFSHMLMDKEEKESMETLRIRRLGEPEDCAGIVSFLCSSEDA SYITGETVW	263
DB	208 vavgfvphfasiflgsssevrreglieekllnrlglgtgmaaaaaafllsdessyltgeltv	264
OY	270 VGGGPSRL 278	
DB	268 vagampstl 276	

RESULT 7

AAG74580

ID AAG74580; standard; Protein: 103 AA.

AC AAG74580;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:5344.

KM Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma.

OS Homo sapiens.

XX WO200112920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

PA (HUMA-) HUMAN GENOME SCI INC.,

XI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX

DR WPI: 2001-235357/24.
 DR N-PSDB; AAH33985.
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 11: Page 6991-6992; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 103 AA;
 Query Match 30.6%; Score 428; DB 22; Length 103;
 Best Local Similarity 94.6%; Pred. No. 2.8e-35;
 Matches 87; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MHAGLLGICARWNSVRNASSGRTDRDPLANKVALVTASTDTCGFAIARLADGAHV 60
 Db 2 mhagllgicaratamsvrnassgmltrdplankvalvtastdgcgfaiaarlgaqdhav 61
 QY 61 VSSRKOONVQAVATLOGEGLSVTGTVCHYK 92
 Db 62 vssrkqgnvdqavatlqgeglsvtgicamwgr 93
 RESULT 8
 AAY54422
 ID AAY54422 standard; Protein: 248 AA.
 AC AAY54422;
 XX
 DT 06-APR-2000 (first entry)
 XX
 DE Amino acid sequence of a beta-ketoacyl-ACP reductase protein.
 XX
 KW Beta-ketoacyl-ACP reductase: Type II fatty acid synthetase;
 KW stereoselectivity; 4-chloroacetoacetic acid ester;
 KW (S)-4-halo-3-hydroxybutyric acid ester; 4-halo-acetoacetic acid ester;
 KW beta-ketoacyl-acyl carrier protein reductase; acetoacetyl-CoA reductase;
 KW polybeta-hydroxy fatty acid biosynthesis; optically active;
 KW 4-haol-3-hydroxybutyric acid ester.
 KW
 OS Bacillus subtilis.
 XX
 PN EP955375-A2.
 XX
 PD 10-NOV-1999.
 XX
 PE 10-MAY-1999; 99EP-0109403.
 XX
 PF 08-MAY-1998; 98JP-0126507.
 PR 21-OCT-1998; 98JP-0300178.
 PR 05-APR-1999; 99JP-0098205.
 XX

PA (DAIL) DAICEL CHEM IND LTD.
 XX
 PI Yamamoto H;
 XX
 DR WPI: 2000-118183/11.
 DR N-PSDB; AA245749.
 XX
 PT Producing optically pure (S)-4-halo-3-hydroxybutyric acid ester -
 XX
 PS Claim 6; Page 19-20; 34pp; English.
 XX
 CC The present sequence represents a beta-ketoacyl-ACP reductase protein
 CC of Bacillus subtilis. The beta-ketoacyl-ACP reductase enzyme constitutes
 CC a type II fatty acid synthetase. The enzyme has an extremely high
 CC reducing activity and stereoselectivity towards 4-chloroacetoacetic
 CC acid ester. The specification describes a method for producing a
 CC (S)-4-halo-3-hydroxybutyric acid ester. The method comprises
 CC asymmetrically reducing 4-halo-acetoacetic acid ester or its
 CC derivative with beta-ketoacyl-acyl carrier protein reductase
 CC constituting Type II fatty acid synthase, or acetoacetyl-CoA
 CC reductase constituting the polybeta-hydroxy fatty acid biosynthesis
 CC system. The novel method is used to produce optically active
 CC 4-haol-3-hydroxybutyric acid ester, with a high purity.
 XX
 SQ Sequence 248 AA;
 Query Match 28.9%; Score 404; DB 21; Length 248;
 Best Local Similarity 37.7%; Pred. No. 2.6e-32;
 Matches 93; Conservative 50; Mismatches 100; Indels 4; Gaps 3;
 QY 28 DPLANKVALVTASTDTCGFAIARLADGAHVVS-SRKOONVQAVATLOGEGLSVTGT 86
 Db 2 dmlndkcalvtgsgvgrlslalalaksaganvvvysgneakanevdelksmgrikaiv 61
 QY 87 VCHVGAKEDEERLVATPAVKLHGIDILVSNAAVNPFGSITMDTTEWMDTIDINKAPA 146
 Db 62 kadvspedvgmnikelvisfstcidilvnmagltr-dnlmrmkedewdvlnlkyvf 120
 QY 147 LMRKAVPEMEKRGSGSVIVSSIAAFSPSPGFSYVNSKTLALGLTKTLAELAPRNR 206
 Db 121 nckkavtrgmkkqgrsilnvssilvysgngpganyaaakayigltksakelaarnt 180
 QY 207 VNCLARGLITFSRMLMNDKEEESKFTLRIRLRGEPDCAGIVSFCSEDASYITGE 266
 Db 181 vnlaipgflstmdtkl--akdvqdemlkqiplarfgepdsvsvtflasegarymtg 238
 QY 267 TVVWGGG 273
 Db 239 tlhidgg 245.
 RESULT 9
 AAG02114
 ID AAG02114 standard; Protein: 79 AA.
 AC AAG02114;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 6195.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 KW
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX

PR 26-FEB-1999; 9905-0122487.
XX
XX (GEST) GENSET.
XX
PI Dumas Mline Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
DR N-PSDB; MAC02120.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 6195; 71pp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs with intact 5'
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
CC
XX Sequence 79 AA:
SO

Query Match 26.4%; Score 370; DB 21; Length 79;
Best Local Similarity 96.2%; Pred. No. 1.3e-29;
Matches 76; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MHKAGLGLICARANMSVMASSGMRDPLANKVALVTASTDGIFFAIRLAODCAHYV 60
Db 1 mxaxxljglcarawsvmassgmrrlplankvalvtastdgi fatarlaodgahyv 60
QY 61 VSSRKQWVDVAATLQGE 79
Db 61 vstrkqnvdaqavallgqe 79

RESULT 10
AAG31818
ID AAG31818 standard; Protein: 262 AA.
XX
AC AAG31818;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 38275.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.

PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0132863.
PR 14-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139452.
PR 18-JUN-1999; 9905-0139453.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142380.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.

[illegible]

	PF	18-DEC--2000; 200DEP-0127688.
	PR	16-DEC-1999; 99JP-0377484.
	XX	07-APR--2000; 200QJP-0159162.
	PR	03-AUG--2000; 200QJP-0280988.
	XX	(KYO) KYOMA HAKKO KOGYO KK.
	PA	Nakagawa S., Mizoguchi H., Ando S., Hayashi M., Ochiai K., Yokoi H;
	XX	Tateishi N., Senoh A., Ikeda M., Ozaki A:
	Pt	WIPI: 2001-376931/40.
	DR	N-PDSB; AAH67301.
	XX	Noval polynucleotides derived from Coryneform bacteria, for identifying
	PT	muation point of a gene, measuring expression of a gene, analysing
	PT	expression profile or pattern of a gene and identifying homologous gene
	PS	-
	XX	Claim 17; SEQ ID NO: 5836; 246bp + Sequence Listing; English.
	CC	The present invention provides a number of nucleotide and protein
	CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These
	CC	are useful for identifying the muation point of a gene derived from a
	CC	mutant of coryneform bacterium, measuring expression amount and
	CC	analysing the expression profile or expression pattern of a gene derived
	CC	from Coryneform bacterium, and identifying a homologue of a gene derived
	CC	from coryneform bacterium. Coryneform bacteria are useful for producing
	CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
	CC	paticularly L-lysine. The present sequence is a protein described
	CC	In the exemplification of the invention.
	CC	Note: The sequence data for this patent did not form part of the printed
	CC	sificication, but was obtained in electronic format directly from the
	EU	European Patent Office.
SQ	Sequence	292 AA:
	Query Match	24.9%; Score 349; DB 22; Length 292;
	Best Local Similarity	34.5%; Pred. No. 1.le-26;
	Matches	86; Conservative 45; Mismatches 114; Indels 4; Gaps 2
OY	DB	30 LANWALVTASTDGIGFAIARRLADGAHYVS--SRQQNDQAVALTDCGGELSVTGTV 87 : : : : : : : : : : : : : : : : : 45 lkykallitgdsqgaavalaalayaagaadavalaylpeeqaadadrilqlgateecgqafsrp 104
OY	DB	88 CHVCAKEERELVAATAVLHGCGIDILVNAAVNPPFGSIMDTEEVMKDITLDINWKAPA 147 : : : : : : : : : : : : : : : : : 105 gdlldpcyrslvygetvnaigldllvnasrqvwaplteitdenfdgtlvnlgysfr 164
OY	DB	148 MKRAVPPEMERGGSGVVIVSSIAFSPSPGFSPYNVKTKALLGLTKTLATLPANRTV 207 : : : : : : : : : : : : : : : : : : 165 vtskaaplhl-kppssillfsslqaygsellldyamtkaalnnlskglassligdgltiv 222
OY	Db	208 NCLAPAGLIKITSFSMLMWDKEEKESMKFTLRIRRGEPEDDAGIVSPLCSDDASTYTGT 267 : : : : : : : : : : : : : : : : : 223 nsvaeprfwrtplrgshhgpgckiegfgqhnaplragnhvvelagalayvfllaadaasyvvge 282
OY	Db	268 VVGGSPTS 276 : : : : : : : : : : : : : : : : : : 283 lyvcgtpct 291
RESULT	12	
ID	AAB79400	standard; Protein: 295 AA.
AC	AAB79400;	
DT	30-Apr-2001	(first entry)
CORYNEBACTERIUM GLUTAMICUM SMP PROTEIN SEQUENCE SEQ ID NO:316.		

KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism; and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogetic amino acid;
 KW nonproteinogetic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100844-A2.
 XX
 PD 04-JAN-2001.
 XX
 XX 23-JUN-2000; 2000WO-IB00943.
 XX
 PF 25-JUN-1999; 99US-0141031.
 XX PR 08-JUL-1999; 99DE-1031412.
 XX PR 08-JUL-1999; 99DE-1031413.
 XX PR 08-JUL-1999; 99DE-1031419.
 XX PR 08-JUL-1999; 99DE-1031420.
 XX PR 08-JUL-1999; 99DE-1031424.
 XX PR 08-JUL-1999; 99DE-1031438.
 XX PR 08-JUL-1999; 99DE-1031431.
 XX PR 08-JUL-1999; 99DE-1031433.
 XX PR 08-JUL-1999; 99DE-1031434.
 XX PR 08-JUL-1999; 99DE-1031510.
 XX PR 08-JUL-1999; 99DE-1031562.
 XX PR 08-JUL-1999; 99DE-1031634.
 XX PR 09-JUL-1999; 99DE-1032180.
 XX PR 09-JUL-1999; 99DE-1032227.
 XX PR 09-JUL-1999; 99DE-1032230.
 XX PR 09-JUL-1999; 99US-0143208.
 XX PR 14-JUL-1999; 99DE-1032924.
 XX PR 14-JUL-1999; 99DE-1032973.
 XX PR 27-AUG-1999; 99DE-1033005.
 XX PR 31-AUG-1999; 99DE-1040765.
 XX PR 03-SEP-1999; 99US-0151572.
 XX PR 03-SEP-1999; 99DE-1042076.
 XX PR 03-SEP-1999; 99DE-1042079.
 XX PR 03-SEP-1999; 99DE-1042086.
 XX PR 03-SEP-1999; 99DE-1042087.
 XX PR 03-SEP-1999; 99DE-1042088.
 XX PR 03-SEP-1999; 99DE-1042095.
 XX PR 03-SEP-1999; 99DE-1042123.
 XX PR 03-SEP-1999; 99DE-1042125.
 XX
 PA (BADI) BASF AG.
 XX
 XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
 DR MPI: 2001-061975/07.
 XX
 DR N-PSDB: AAF71517.
 XX
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 PS Claim 20; Page 585-586; 1246pp; English.
 XX
 XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (11) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogetic
 CC or nonproteinogetic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (i) or SMP proteins
 CC (11i) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells

PR 08-JUL-1999; 99DE-1031562.

AAE71366 to AAE71350 encode the Corynebacterium glutamicum sugar
metabolism and oxidative phosphorylation (SMP) proteins given in
AAE79243 to AAE 79633 which are involved in carbon metabolism and
energy production. The C. glutamicum SMP gene can be used in vectors
(II) for expression in host cells and production or modulation of
production of fine chemicals, such as, an organic acid, a proteino-genic
or nonproteino-genic amino acid (preferred), a purine or pyrimidine base,
a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
(III) encoded by them are used for diagnosing the presence or activity of
Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
containing them are used to map genomes of organisms related to
C. glutamicum, identify and localise C. glutamicum sequences of interest,
in evolutionary studies, in determining SMP protein regions required
for function, in modulating SMP protein activity, in modulating the
metabolism of sugars, and in modulating high-energy molecule production
in a cell (i.e. ATP, NADPH).

[illegible]

QY 33 KVALVTASIDIGIFALARRLADGAAHVVS-SRKQONVDQAVATLOEGSLSVTGFCHVG 91
| |||| :: ||| :: | : : : : | : : : :
Db 5 ksalvtgsgrgirgsialqlaeegynvavnyagsrkeaaevveeikaygvdsfaiqanva 64

biological agents affecting fatty acid biosynthesis, compounds
contracting a bacterial enzymatic pathway with enzymes. The method is
used for screening biological agents affecting fatty acid biosynthesis.
Agonists and antagonists of fat (fatty acid biosynthesis) are used to
inhibit, prevent or treat diseases such as infections of the upper
respiratory tract (e.g. otitis media, bacterial tracheitis, acute
epiglottitis, thyroiditis), lower respiratory (e.g. emphysema, lung
abscess), cardiac (e.g. infective endocarditis), gastrointestinal (e.g.
secretory diarrhoea, splenic abscess, retroperitoneal abscess), central

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 08:16:31 ; Search time 16.41 Seconds

(without alignments)
1290.465 Million cell updates/sec

Title: US-09-866-034-2

1399

Sequence: 1 MHKAGLGLGCAKRAMSVNMA.....DASYITGETIVVGGTIPSRL 278

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR.68:**
1: PIR1:**
2: PIR2:**
3: PIR3:**
4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	824.5	58.9	280	2	S6665	nuclear protein H
2	636	45.5	257	2	T32002	hypothetical prot
3	614	43.9	260	2	T22676	hypothetical prot
4	588.5	42.1	254	2	E85069	hypothetical prot
5	459.5	32.8	260	2	E70881	probable fabG5 pro
6	409	29.2	255	2	D83416	probable short-cha
7	399	28.5	246	2	A69621	3-oxoacyl-(acyl-ca
8	396.5	28.3	246	2	H72219	3-oxoacyl-(acyl-ca
9	388	27.7	246	2	C83961	3-oxoacyl-(acyl-ca
10	385	27.5	248	2	H70447	3-oxoacyl-(acyl-ca
11	381.5	27.3	282	2	C71204	probable dehydrog
12	381	27.2	255	2	H84288	3-oxoacyl-(acyl-ca
13	374	26.7	243	2	G72389	hypothetical prote
14	373.5	26.7	251	2	F86721	oxidoreductase, s
15	371	26.5	263	2	C75217	probable short-cha
16	370	26.4	249	2	B83767	glucose 1-dehydrog
17	369.5	26.4	285	2	D85885	probable oxidoredu
18	368.5	26.3	285	1	A65017	3-oxoacyl-(acyl-ca
19	368.5	26.3	320	2	S22450	3-oxoacyl-(acyl-ca
20	366	26.2	261	2	S35196	hypothetical prote
21	365.5	26.1	248	2	F69688	glucose 1-dehydrog
22	364.5	26.1	260	2	B48674	troponin reductas
23	364	26.0	234	2	G75333	troponin reductas
24	360.5	25.8	261	2	S54815	urdamycin polyket
25	360	25.7	272	1	S05397	gramicidin polyket
26	358.5	25.6	256	2	A70637	hypothetical prote
27	357	25.5	248	2	F81971	probable 3-oxoacyl
28	357	25.5	248	2	E81026	3-oxoacyl-(acyl-ca
29	357	25.5	273	1	A48674	troponin reductas

30	354	25.3	241	2	F71636	3-oxoacyl reductas
31	354	25.3	278	2	T10877	y41A protein - Rhl
32	353	25.2	262	2	B84695	probable troponin
33	349	24.9	256	2	E72427	oxidoreductase, sh
34	349	24.9	258	2	G69755	glucose 1-dehydrog
35	347.5	24.8	253	2	E83207	probable short-cha
36	345.5	24.7	268	2	B84693	probable troponin
37	345	24.7	246	1	DEK6NG	acetoacetyl-CoA re
38	345	24.7	248	2	D70707	hypothetical prote
39	344.5	24.6	255	2	G82644	2,5-dichloro-2,5-c
40	343	24.5	248	2	F82128	3-oxoacyl-(acyl-ca
41	342	24.4	247	2	S77280	3-oxoacyl-(acyl-ca
42	341.5	24.4	246	2	H84136	3-oxoacyl-(acyl-ca
43	341	24.4	248	2	E81695	probable troponin
44	341	24.4	268	2	E84695	2-deoxy-D-gluconat
45	341	24.4	281	2	F69400	

ALIGNMENTS

```
RESULT 1
S6665
nuclear protein Hep27 - human
N:Alternate names: protein D
C:Species: Homo sapiens (man)
C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: S6665; S66614; S14094; S78122
R:Gabrielli, F.; Donadel, G.; Bensli, G.; Heguy, A.; Melli, M.
A:Title: A nuclear protein, synthesized in growth-arrested human hepatoblastoma cell.
A:Reference number: S66614; MUID:96035881
A:Accession: S6665
A:Molecule type: mRNA
A:Residues: 1-280 <GAB>
A:Cross-references: EMBL:U31875; NID:G1079565; PID:AA82048.1; PID:G1079566
A>Note: Met-23 is the probable initiator
A:Accession: S66614
A:Molecule type: protein
A:Residues: 24-41 <GAB>
R:Donadel, G.; Gabrielli, C.; Frank, R.; Gabrielli, F.
Eur. J. Biochem. 195, 723-729, 1991
A:Title: Identification of a novel nuclear protein synthesized in growth-arrested hu
A:Reference number: S14094; MUID:91153312
A:Accession: S14094
A:Molecule type: protein
A:Residues: 80-83, 'L', 85, 87-88, 'A', 90-91, 141-146, 148-162, 198-205 <DON>
A:Accession: S78122
A:Molecule type: protein
A:Residues: 'V', 81-84, 'G', 86-88 <DOM>
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: nucleus
F:37-216/Domain: short-chain alcohol dehydrogenase homology <SABH>

Query Match 58.9% Score 824.5; DB 2; Length 280;
Best Local Similarity 61.6% Pred. No. 3.4e-58;
Matches 162; Conservative 47; Mismatches 51; Indels 3; Gaps 2;

OY 10 CARAMNSVRRMSSMTTRDPLANKVALVYASTDGIQFARLRLODGAHVVSRRKQONV 69
16 CARL--SVRRMSTSIDRKGVLANVAVVGTSGIGFAIRLRARDGAHVVISRRKQONV 73
OY 70 DOAVATLOGEGLSVTGVCHVCHGKAEDREBLVATVAKLHGIDILVNAANVPFGSTMDV 129
Db 74 DRAMAKLOGEGLSVAGVCHVCHGKAEDREBLVAKLHGCGGVDFVCSAGVNPVLTGTLGT 133
OY 130 TEVYMDKTLIDINKAPALMKRAVPEMEKRGCGSVTVSSIAFSPSPGSPVNSKTAL 189
Db 134 SEQIWDKILSVNWKSPALLLSQLPYMENR--RGAVILVSSIAVNPVVALGVTVNSKTAL 192
OY 190 LGITKTLATLAPRNIVNCLAPGLITSTFSRLMDKEESNKKETLRIRLGEPEDECA 249
```

[illegible]

A:Accession: T24959
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-260 <MI2>
A:Cross-references: EMBL:281592; PTDN:CAB04734.1; GSPDB:GN00023; CESP:F54F3.4
A:Experimental source: clone T1601
C:Genetics:
A:Gene: CESP:F54F3.4
A:Map position: 5
A:Introns: 21/2; 81/3; 201/3
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 43.9%; Score 614; DB 2; Length 260;
Best Local Similarity 48.0%; Pred. NO. 1.6e-41;
Matches 120; Conservative 56; Mismatches 70; Indels 4; Gaps 2;

QY 33 KAAIVTAASDGIIGFATARRLADGAAHYVYSSKKQONDOAVATLOGEGLS-VTGTVCHVG 91
Db 11 KAAIVTAARKGIGLAEHLDEGASVYVSSNOKNVEALIEYLKGLTKVAGCIAGHIA 70

QY 92 KAEDERLVAIVKLLHGGLDILVNAAVPFGSLMDYTEEYWDKTLIDINAKPALMKA 151
Db 71 SYDDOKKIVLDFLQKFGKINTLVNHHGIMPATGHLLEVSQVWDKLFEEVNAKAGQMTKL 130

QY 152 VPEMEKRGGSVVIVSSIAAFSPGSPYVNSTALLGLTKTALAEAPRNIRVNCIA 211
Db 131 VHPHIAKEGGAGIIFNAYSAYKSPPGIAAYGVTKTLVLGLRALAMGLANDINRVNGIA 190

QY 212 PGLITSTPSRMV---MKEKEESMKELIRLRRLCEPEDECACTSYFLCSEASVYTGSETV 268
Db 191 PGLVITKMSQVIMDGEDAEKELTDIOETALGLRGPDDCAGTAVAYLASDDSSYITGEMI 250

QY 269 VVGGGTPSRLL 278
Db 251 ILAGVQARL 260

RESULT 4
E85069
hypothetical protein AT4g05530 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: E85069
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: E85069
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:Cross-references: GB:NC_001268; NID:g7267313; PTDN:CAB81095.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g05530
A:Map position: 4

Query Match 42.1%; Score 588.5; DB 2; Length 254;
Best Local Similarity 50.6%; Pred. NO. 1.7e-39;
Matches 126; Conservative 33; Mismatches 85; Indels 3; Gaps 2;

QY 30 LANNVALVASTDGIIGFAIARRLAODGAHYVYSSRKQONDOAVATLOGEGLSVTGVCH 89
Db 9 LEGKVAIVTASTGIGFGITERFGLGASVYVSSRKQANDENAVAKLKSIGDAIVGIVCH 68

QY 90 VGRADEDERLVAIVKLLHGGLDILVNAAVPFGSLMDYTEEYWDKTLIDINAKPALMT 149
Db 69 VSNQOHRNRLIVETKRYOKYKIDIVVCNAANPSTDPILSKSEAVLDKLTMEINVKSSITLL 128

QY 150 KAAVPEMEKRGGSVVIVSSIAAFSPGSPYVNSTALLGLTKTALAEAPRNIRVNC 209
Db 129 QDMAVPLEK--GSSVIVITTTIAGFSFGCAAMAGVTTALLGLTKALAAEMAP--DFRVNA 185

RESULT 7
A69621
3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) [validated] - *Bacillus subtilis*
N:Alternate names: 3-ketoacyl-acyl carrier protein reductase
C:Species: *Bacillus subtilis*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C:Accession: A69621, PC4176, I46633
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ben-
C. E.; Bron, S.; Brüllert, S.; Brusch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foidger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Gallizi, A.; Gal-
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, P.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidas, A.; Larin-
A:Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma-
Y., M.; Ogawa, K.; Ogizawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet-
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadie, Y.; Sato, T.; Scan-
A:Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se-
A:Authors: Winters, P.; Wipit, A.; Tanaka, T.; Terpiltra, P.; Tognoni, A.; Tosato, V.; Uchi-
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
A:Reference number: A69580; MUID:98044033
A:Accession: A69621
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-246 <KUN>
A:Cross-references: GB:599112; GB:AL009126; NID:g2633902; PIDN:CAB13464.1; PID:g2633337012
A:Experimental source: strain 168
R:Okuro, A.; Kakeshita, H.; Takamitsu, H.; Nakamura, K.; Yamane, K.
Gene 172, 17-24, 1996
A:Title: The effect of Srb, a homologue of the mammalian SRP receptor alpha-subunit,
A:Reference number: JG4819; MUID:96257247
A:Accession: PC4176
A:Molecule type: DNA
A:Residues: 230-246 <OGU>
A:Cross-references: DDBJ:064116; NID:g1389548; PIDN:BAI0974.1; PID:g1237012

Query Match 27.3%; Score 381.5; DB 2; Length 282;
Best Local Similarity 37.7%; Pred. No. 5.2e-23;
Matches 98; Conservative 49; Mismatches 100; Indels 13; Gaps 6

```

RESULT 13      :
F86721         :
hypothetical protein fabG1 (imported) - Lactococcus lactis subsp. lactis (strain IL1)

```

D

QY 33 6	KVALYSTOIGCFIARLRLODGAHV---VSSRKOONDOAVATLLOGEGSVTGTVC89 :: : :: : KVLTLTGAGSGICGRKAAMFAERGAARVAINDISEEKGKETVELIKSMGEAFAIFGDB--63
-------------------	--

Search completed: May 8, 2002, 08:24:40
Job time: 489 sec

OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98196666; PubMed=9537320;
RC STRAIN=VF5.
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Grahm D.E., Overbeek R., Sned M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:355-358(1998).
CC -I- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
CC -I- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SRR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000752; AAC07575.1; -
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR002347; Adh_short_C2.
DR Pfam; PF00106; adh_short_1.1
DR Pfam; PF00678; adh_short_C2; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PRINTS; PR00081; GDRDH.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Fatty acid biosynthesis; Oxidoreductase; NADP: Complete proteome.
FT NP_BIND 12 36 NADP (BY SIMILARITY).
FT ACT_SITE 157 157 BY SIMILARITY.
SQ SEQUENCE 248 AA; 26867 MW; 3CFDEB9AD83F2C5 CRC64;

OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98196666; PubMed=9537320;
RC STRAIN=VF5.
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Grahman D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:355-358(1998).
CC -I- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
CC -I- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SCR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000752; AAC07575.1;
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR002347; Adh_short_C2.
DR Pfam; PF00106; adh_short_1.1
DR Pfam; PF00678; adh_short_C2; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PRINTS; PR00081; GDRDH.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Fatty acid biosynthesis; Oxidoreductase; NADP: Complete proteome.
FT NP_BIND 12 36 NADP (BY SIMILARITY).
FT ACT_SITE 157 157 BY SIMILARITY.
SQ SEQUENCE 248 AA; 26867 MW; 3CFDEB9AD83F2C5 CRC64;

Query Match	27.1%	Score 379.5;	DB 1;	Length 260;
Best Local Similarity	35.4%	Pred. No. 2.8e-21;		
Matches 90; Conservative	49;	Mismatches 110;	Indels 5;	Gaps 3.

Db
1 MAGRWNEGCTALVTGSGRGICYGVIELANLGASVYTCSSRNQCKELDECTLTQMRSGCNV 60

between Bioinformatics and the EMBL outstation -
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC

Query Match	26.7%	Score 373.5	DB 1	Length 251
Best Local Similarity	36.5%	Pred. No. 7.4e-21		
Matches 91	Conservative 52	Mismatches 93	Indels 13	Gaps 5

[illegible]

ID	TRN2	DATST	STANDARD	PRT	260 AA
Db	52	IATSGIAQVATAAKKASAGACQSVESPVLYVTGASRGTCAIALSLACKACQKVLVYNARS	111		
Qy	63	SRKOONDOAVATLOGESLVTGCVHGAEDREBRVATAVAKLHGIDILVSNAAVNPF	122		
Db	112	SKKEEVSKEIEAFGGALTFEG---DYSKEEDYEAIKTAIVAMGIVDLVNNAGITR-	167		
Qy	123	FGSIMDTEEMDKTLDINKARPLMKAVPEMEKRGSSVIVSSIAAFSPSPGFSPI	182		
Db	168	DGLIMRRKKQWOMEVDLNLGTGFLCTQAAKIMKKKGRRIINIASVGVAGNAGANY	227		
Qy	183	NVSKATLGLTTLTALTELAPRINRVNCLAPGLIKTSRRLAMDKEEBSKETLRRL	242		
Db	228	SAKAGVIGTGTVAKRYASRNINVAAGFISDDMTSKLGDINKK--ILFTIPIGRY	285		
Qy	243	GEPEDCAGIVSFLCSEDA-SYITGETVWVGW 273			
Db	286	GOPEEVAGIVEFLAINPASSVYGQVFTIDGG 317			
RESULT	8				
TRN2	DATST	STANDARD	PRT	260 AA	
AC	P50163	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	TROPINONE REDUCTASE-II (EC 1.1.1.236) (Trn-II).				
GN	TR2.				
OS	Datura stramonium (Jimsonweed) (Common thornapple).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Asteridae; eusteriids I; Solanales; Solanaceae; Datura.				
OX	NCBI_TaxID=4076;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=ROOT:				
RA	MEDLINE=94022421; PubMed=8415746;				
RA	Nakajima K., Hashimoto T., Yamada Y.;				
RT	"Two tropinone reductases with different stereospecificities are				
RT	short-chain dehydrogenases evolved from a common ancestor.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:9591-9595(1993).				
RN	[2]				
RP	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).				
RA	MEDLINE=98226735; PubMed=9560196;				
RA	Nakajima K., Yamashita A., Akama H., Nakatsu T., Kato H.,				
RA	Hashimoto T., Oda J., Yamada Y.;				
RT	"Crystal structures of two tropinone reductases: different reaction				
RT	stereospecificities in the same protein fold.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:4876-4881(1998).				
RN	[3]				
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).				
RA	MEDLINE=99316165; PubMed=10387002;				
RA	Yamashita A., Kato H., Nakatsuki S., Tomizaki T., Nakatsu T.,				
RA	Nakajima K., Hashimoto T., Yamada Y., Oda J.;				
RT	"Structure of tropinone reductase-II complexed with NADP+ and				
RT	pseudotropine at 1.9 A resolution: implication for stereospecific				
RL	substrate binding and catalysis.";				
RL	Biochemistry 38:7630-7637(1999).				
CC	-1- FUNCTION: CATALYZES THE STEREOSPECIFIC REDUCTION OF TROPINONE TO				
CC	PSEUDOTROPINE.				
CC	-1- CATALYTIC ACTIVITY: PSEUDOTROPINE + NADP(+) = TROPINONE + NADPH.				
CC	-1- PATHWAY: BIOSYNTHETIC PATHWAY OF TROPANE ALKALOIDS.				
CC	-1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES				
CC	(SDR) FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				

```

CC  EMBL; L20474; AAA33282.1;
DR  PDB; 2AE2; 18-NOV-98.
DR  PDB; 2AE2; 02-FEB-99.
DR  InterPro; IPR002198; ADH_short.
DR  InterPro; IPR002347; Adh_short_C2.
DR  Pfam; PF00106; adh_short_1.
DR  Pfam; PF00678; adh_short_C2; 1.
DR  PRINTS; PR00080; SDRFAMILY.
DR  PRINTS; PR00081; GDHRDH.
DR  PROSITE; PS00061; ADH_SHORT; 1.
KW  Oxidoreductase; NADP: 3D-structure.
FT  NP_BIND 13 37 NADP (BY SIMILARITY).
FT  ACT_SITE 159 159 BY SIMILARITY.
SQ  SEQUENCE 260 AA; 28311 MW; 2DBF4963B2CACA303 CRC64;

```

```

Query Match          26.1%  Score 364.5; DB 1; Length 260;
Best Local Similarity 34.6%  Pred. No. 3.6e-20;
Matches 88; Conservative 50; Mismatches 111; Indels 5; Gaps 3;

```

```

QY  24 MTRRDLANKVALVTASTDGIQFAIRRLAQDGAHVYSSRKOONVDAVATLOGEGSLV 83
DB  1 MAGRWNLGECTALVTGSGSIGIYVIELASLGASVYTCGRNCKELDTQNRSGKFV 60
QY  84 TGVVCHVGAEDRELVAT-AVKLHGIDILVSNAAVNPFFGSDIMVTEWMDKTIDIN 142
DB  61 EASVCLSSRSEKQELMNTVANHFGKLNILVNNAGI-VYKAKQYVDEYSLINSIN 119
QY  143 KAPALMTKAVPEMEKRGGSVIVASSIAFSPSPSPINVKTLGLTKTLAIELAP 202
DB  120 EAAVHLSTVAHPFLKASERNVAFISSGALAVPEAVGATKGMADQTRCLAEEMAK 179
QY  203 RIRIRVCLAPGLIKTSFSLMMDKEKSEMKETL--RIRRLGPEDDAGIYFICSED 259
DB  180 DNIRVNGVGPVIATVSLVEMTIDPQKKNLKLIRCALRRNGEKEFLAAMVAFICFPA 239
QY  260 ASYITGETVYVGGG 273
DB  240 ASYVTGQITLYVDGG 253

```

```

RESULT 9
UCPA_ECOLI          STANDARD; PRT; 263 AA.
AC  P37440; P77442; P76963; P77140;
DT  01-OCT-1994 (Rel. 30, Created)
DT  20-AUG-2001 (Rel. 40, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  OXIDOREDUCTASE UCPA (EC 1.-.-.-).
GN  UCPA OR B2426.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX  NCBI_TaxID=562;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MG1655;
RX  MEDLINE=97426617; PubMed=9278503;
RA  Blattner F.R., Plunkett G. III, Bloch C.A., Petina N.T., Burland V.,
RA  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Mau B., Shao Y.;
RT  "The complete genome sequence of Escherichia coli K-12."
RL  Science 277:1453-1474(1997).
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12;
RX  MEDLINE=97349980; PubMed=9205837;
RA  Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA  Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA  Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA  Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,

```

```

RA  Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA  Yamagata S., Horiuchi T.;
RT  "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT  K-12 genome corresponding to 50,0-68.8 min on the linkage map and
RT  analysis of its sequence features."
RL  DNA Res. 4:91-113(1997).
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12;
RX  MEDLINE=97385354; PubMed=9241368;
RA  Sirko A., Weglenska A., Hryniewicz M.M., Hulanicka D.M.;
RT  "Characterization of the Escherichia coli gene encoding a new member
RT  of the short-chain dehydrogenase/reductase (SDR) family."
RL  Acta Biochim. Pol. 44:153-157(1997).
RP  SEQUENCE OF 180-263 FROM N.A.
RC  STRAIN=K12;
RX  MEDLINE=90264335; PubMed=2188959;
RA  Hryniewicz M.M., Sirko A., Palucha A., Boeck A., Hulanicka D.M.;
RT  "Sulfate and thiosulfate transport in Escherichia coli K-12:
RT  identification of a gene encoding a novel protein involved in
RT  thiosulfate binding."
RL  J. Bacteriol. 172:3358-3366(1990).
RN  [5]
RP  IDENTIFICATION.
RX  MEDLINE=95075659; PubMed=7984428;
RA  Borodovsky M., Rudd K.E., Koonin E.V.;
RT  "Intrinsic and extrinsic approaches for detecting genes in a
RT  bacterial genome."
RL  Nucleic Acids Res. 22:4756-4767(1994).
CC  -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC  (SDR) FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL; AE000330; AAC75479.1; ALT_INIT.
DR  EMBL; D90872; BAAL6309.1; ALT_INIT.
DR  EMBL; D90871; BAA16300.1;
DR  EMBL; X99908; CAA68181.1;
DR  EMBL; M32101; -, NOT_ANNOTATED_CDS.
DR  HSSP; P14061; IFDV.
DR  Ecolene; EG12133; ucpa.
DR  InterPro; IPR002198; ADH_short.
DR  InterPro; IPR002347; Adh_short_C2.
DR  Pfam; PF00106; adh_short_1.
DR  Pfam; PF00678; adh_short_C2; 1.
DR  PRINTS; PR00080; SDRFAMILY.
DR  PROSITE; PS00061; ADH_SHORT; FALSE_NEG.
KW  Oxidoreductase; Complete proteome.
FT  CONFLICT 153 153 T->Q (IN REF. 3)
FT  SEQUENCE 263 AA; 27850 MW; 2E201713357FDF41 CRC64;

```

```

Query Match          25.9%  Score 362.5; DB 1; Length 263;
Best Local Similarity 34.1%  Pred. No. 5.1e-20;
Matches 86; Conservative 53; Mismatches 104; Indels 9; Gaps 4;
QY  30 LANKVALVTASDGIQFAIRRLAQDGAHVYSSRKOONVDAVATLOGEGSLVGTGCH 89
DB  4 LTGKTAALTGALGIGEGIAFTFARGANLILDISPE-IEKADLGLGRGRTAVVAD 62
QY  90 VKAADRELVATFAVKLHGIDILVSNAAVNPFFGSDIMVTEWMDKTIDINVKAPALMT 149
DB  63 VQDPASVAALIRAKEREGRIDILVNNAGVCR-LGSEFLDMSDDDRFDHIDINIKGVNWT 121
QY  150 KAVPEMEKRGGSVIVASSIA-FSPSPGSPYNNVSKTALLGLTKTLAIELAPRIRIN 208

```

```

DB 122 KALPELMARAKDRIYMWSSVYTGDMVADPGETAYALTKAIVGLTKSLAVEYAOQSGIRVN 181
QY 209 CLAPGILKTSFSLMWNDEKEE-----SMKFTLRIRRGEPEDCAGIYSLCSEBASY 262
DB 182 AICPGVIRTPMAESTAROSNPEDPESVLTETMAKAIIPRLADPLEVGEALFASDESSY 241
QY 263 TGETVWVGCGT 274
DB 242 LIGTONVIDGGS 253

RESULT 10
DHKL_STRVN STANDARD: PRT: 272 AA.
AC P16542:
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE GRANATICIN POLYKETIDE SYNTHASE PUTATIVE KETOACYL REDUCTASE 1
DE (EC 1.3.1.-) (ORF5).
OS Streptomyces violaceoruber.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1935;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TU22:
RX MEDLINE=90060034; PubMed=2583128;
RA Sherman D.H., Maipartida F., Bibb M.J., Kieser H.M., Bibb M.J.,
RA Hopwood D.A.;
RT "Structure and deduced function of the granaticin-producing
RT polyketide synthase gene cluster of Streptomyces violaceoruber
RT Tu22."
RL EMO J. 8:2717-2725(1989).
CC -1- PATHWAY: BIOSYNTHESIS OF THE POLYKETIDE ANTIBIOTIC GRANATICIN.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X16300; CAA34368.1; -
DR EMBL: X16144; CAA34263.1; -
DR PIR: S05397; S05397.
DR HSSP: P03617; 1IFK.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_1.
DR Pfam: PF00678; adh_short_C2_1.
DR PRINTS: PR00080; SDRFAMILY.
DR PRINTS: PR00081; GDRDH.
DR PROSITE: PS00061; ADH_SHORT_1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT NP_BIND 21 45 NAD (BY SIMILARITY).
FT ACT_SITE 168 168 BY SIMILARITY.
SO SEQUENCE 272 AA; 28393 MW; 0DE54415A19FB218 CRC64;

Query Match 25.7%; Score 360; DB 1; Length 272;
Best Local Similarity 34.8%; Pred. No. 8.1e-20;
Matches 92; Conservative 45; Mismatches 95; Indels 32; Gaps 6;
QY 31 ANKVALVASTDGIQFATARLADGCAHVYSSRKQANDQAVALLQEGSLSYTGVCHV 90
DB 16 AKPVALVGAISGIGLAFARLALGAFETLCARDEERLAQTVKELRGSGFDVGTCDV 75
QY 91 GRAEDRERLVATAVAKHGIDILVSNAAVNPFPGSIMDTEFVMDKTLIDINKAPALMTK 150

```

```

DB 76 ADPAQIRAYVAAAQRGYDILVNNNG-RSGGATPEIDELMDVITTNLTVPLMTK 134
QY 151 AYVPE-----MEKRG-----GGSVIVSSIAFSPSPSPYNNKFTALLGLFTLAI 198
DB 135 EVLNAGCMALKRRGRININASTGKGQGVHAV-----PYASKHGCVGLTALGL 184
QY 199 ELAPRNIRVNCIAPGLTKTSFSR-----LWMDKKESEKTELRI--RIUGPEDCA 249
DB 185 ELARTGITVNAVCGPFEVTPMAERVRHYAGIMOVSEETFRITNVPLGRVYETREVA 244
QY 250 GIVSFLCEDASYTGTGVVCGG 273
DB 245 AMVEYLVAADDAVTAQALVNCGG 268

RESULT 11
TRNL_DATST STANDARD: PRT: 273 AA.
AC P50162:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TROPINONE REDUCTASE-I (EC 1.1.1.206) (TR-I) (TROPINE DEHYDROGENASE).
OS Datura stramonium (Jimsonweed) (Common thornapple).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Datura.
OX NCBI_TaxID=4076;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-ROOT;
RX MEDLINE=94022421; PubMed=8415746;
RA Nakajima K., Hashimoto T., Yamada Y.;
RT "Two tropinone reductases with different stereospecificities are
RT short-chain dehydrogenases evolved from a common ancestor."
RL Proc. Natl. Acad. Sci. U.S.A. 90:9591-9595(1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=98226735; PubMed=9560196;
RA Nakajima K., Yamashita A., Akama H., Nakatsu T., Kato H.,
RA Hashimoto T., Oda J., Yamada Y.;
RT "Crystal structures of two tropinone reductases: different reaction
RT stereospecificities in the same protein fold."
RL Proc. Natl. Acad. Sci. U.S.A. 95:4876-4881(1998).
CC -1- FUNCTION: CATALYZES THE STEREOSPECIFIC REDUCTION OF TROPINONE TO
CC TROPINE.
CC -1- CATALYTIC ACTIVITY: TROPINE + NADP(+) = TROPINONE + NADPH.
CC -1- PATHWAY: BIOSYNTHETIC PATHWAY OF TROPANE ALKALOIDS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L20473; AAA33281.1; -
DR PDB: 1AE1; 18-NOV-98.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_1.
DR Pfam: PF00678; adh_short_C2_1.
DR PRINTS: PR00081; GDRDH.
DR PROSITE: PS00061; ADH_SHORT_1.
KW Oxidoreductase; NADP; 3D-structure.
FT NP_BIND 25 49 NADP (BY SIMILARITY).
FT ACT_SITE 171 171 BY SIMILARITY.

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: AE000082; AAB91754.1; -
 DR HSSP: P29132; 1DEF1.
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2_1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PRINTS: PR00081; GDRHD.
 DR PROSITE: PS00061; ADH_SHORT_1.
 DR Hypothetical protein: Oxidoreductase: Plasmid.
 FT NP_BIND 10 34 NAD OR NADP (BY SIMILARITY).
 FT ACT_SITE 159 159
 FT SEQUENCE 278 AA; 28743 MW; 1D0105625BE9DF2B CRC64;

Query Match 25.3%; Score 354; DB 1; Length 278;
 Best Local Similarity 35.5%; Pred. No. 2.3e-19;
 Matches 87; Conservative 46; Mismatches 106; Indels 6; Gaps 3;

QY 33 KVALVTASTDGIIGFAIARRLAODGAHVYSSRKQNVDAVATLOGEGLSTVGTCHVK 92
 DB 7 KVAAYTAGAGIGKACALAINREGGRVYADLDGSAIACIAEAGNALAMADIND 66
 QY 93 AEDREIVATVAKLHGIDILVSNAA---VNEFGSINDVTEEVNDKTLIDINAKPALMT 149
 DB 67 AQAVALLETETERNHGVLDLVNNAAMHLPDRALIDLDLAVMDQTMATNLRGTLCC 126
 QY 150 KAVPEMEKRGGSVYVISTIAFSPGFPSPYNSKALLGLTTLAIELAPRIRVNC 209
 DB 127 KQATPRMARGGGATVNSGCGLSGDTAQTSTYAVSKAAMNLSASLTQYGHQAIRCNA 186
 QY 210 LAPGLIKTSFSRML-WMDKEEESMKETLRIRLGEPEDCAGIYFLCSEDSYTGTV 268
 DB 187 VAPGLIMT-ERLLAKLDKCMQRHLRSHQLLPRVGHPEVDALVAFLLSDDSFTGOVY 244
 QY 269 VVGCG 273
 DB 245 CIDGG 249

RESULT 14
 ID Y019_THEME STANDARD; PRT; 256 AA.
 AC O56318;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PUTATIVE OXIDOREDUCTASE TM0019 (EC 1.1.1.17).
 GN TM0019.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;
 RX MEDLINE=96125254; PubMed=8550425;
 RA Kletzin A., Adams M.,
 RT "Molecular and phylogenetic characterization of pyruvate and 2-
 RT ketoglutarate ferredoxin oxidoreductases from *Pyrococcus furiosus*
 RT and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*,"
 RL J. Bacteriol. 178:248-257(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;

RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linder K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Salzberg S.L., Smith G.G., Fleischmann R.D., Eisen J.A., White O.,
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of *Thermotoga maritima*,"
 RL Nature 399:323-329(1999).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: X85171; CAAS9459.1; -
 DR EMBL: AE001690; AAD35113.1; -
 DR HSSP: P19992; 2HSD.
 DR TIGR: TM0019;
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2_1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PRINTS: PR00081; GDRHD.
 DR PROSITE: PS00061; ADH_SHORT_1.
 DR Hypothetical protein: Oxidoreductase: NADP; Complete proteome.
 FT NP_BIND 9 33 NADP (BY SIMILARITY).
 FT ACT_SITE 153 153
 FT CONFLICT 130 136 RGGVLI -> TRNRSQH (IN REF. 1).
 FT SEQUENCE 256 AA; 28078 MW; D68160B1D7980C6B CRC64;

Query Match 24.9%; Score 349; DB 1; Length 256;
 Best Local Similarity 35.2%; Pred. No. 4.9e-19;
 Matches 89; Conservative 45; Mismatches 103; Indels 16; Gaps 5;

QY 30 LANKVALVTASTDGIIGFAIARRLAODGAHVYSSRKQNVDAVATLOGEGLSTVGTCH 89
 DB 2 LEGKVAAYTGGGCGIGAIQLFAENCKMVAIAIDEAGVEREMLERGLDVFYKTD 61
 QY 90 VKAEDREIVATVAKLHGIDILVSNAAVNEFGSINDVTEEVNDKTLIDINAKPALMT 149
 DB 62 VADENSVAKNMRYKVEITGVADVNNNAVMS-VKSIERPLEEERIVRLNLTGPYCS 120
 QY 150 KAVPEMEKRGGSVYVISTIAFSPGFPSPYNSKALLGLTTLAIELAPRIRVNC 209
 DB 121 RYCAEMIKRGGSVYVISTIAFSPGFPSPYNSKALLGLTTLAIELAPRIRVNC 180
 QY 210 LAPGLIKTSFSRML-WMDKEEESMKETLR-----IRLGEPEDCAGIYFLC-SEDA 260
 DB 181 ISPGWIETSS---EW--KKSLRKKRDLRPIDHQHPAGRGVGNFLDLAHLVFLADDEKA 234
 QY 261 SYITGTVVVGCG 273
 DB 235 GFITGTNEIVDGG 247

RESULT 15
 ID DHG2_BACSU STANDARD; PRT; 258 AA.
 AC P80869;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GLUCOSE 1-DEHYDROGENASE II (EC 1.1.1.47) (GLCDH-II) (GDH-II) (GENERAL

```

DE STRESS PROTEIN 74) (GSP74).
GN YCDF.
OS Bacillus subtilis.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=168.
RA Kunano M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation.";
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
RN [2].
RC SEQUENCE OF 1-13.
RC STRAIN=IS58;
RX MEDLINE=97443988; Pubmed=9298659;
RA Anselmann H., Bernhard J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
-1 CATALYTIC ACTIVITY: BETA-D-GLUCOSE + NAD(P)(+) = D-GLUCONO-DELTA-
LACTONE + NAD(P)H.
-1 SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
-1 INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE
LIMITATION AND OXYGEN LIMITATION.
-1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB000617; BAA2244.1; -
DR EMBL: Z99105; CAB12077.1; -.
DR SUDILast; BG12761; YCDF.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR002347; Adh_short_C2.
DR Pfam; PF00106; adh_short.1.
DR Pfam; PF00678; adh_short_C2; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PRINTS: PR00081; GDBRDI.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NADP; Complete proteome.
FT NP_BIND 11 35 NADP (BY SIMILARITY).
FT ACT_SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 258 AA; 27776 MW; 77842DC45D96C26 CRC64;

```

Query Match	24.9%	Score 349	DB 1	Length 258
Best Local Similarity	33.9%	Pred. No. 4.9e-19		
Matches	87	Conservative 53	Mismatches 105	Indels 12
			Gaps	5
QY	30	LANKALVTASTDGIGFGLARRLRADGGHHVYSSRKO-QNQDAVALVLOGEGLSVTGTVC	88	
Db	5	LTGKAIVATYSGSKGKGLIAERFGCEKKNNVYVNHSDSGADELTLEIKONGKRAVSYEA	64	
QY	89	HVGAKEDEERLVATAVKLHGIGIDILVNSAANVPFGSIIMDYEEVWDKTLINVKAPALM	148	
Db	65	DVSKEEGIGIALLDLTHFFGTLIDVWVNNSSGNGCYAMHEHMSLEDMQWIRIDVNVGTFLG	124	
QY	149	TKAAVPEKEKRG-GGSVVIVSSIAAFSPGSPRYNYSKTLGLGTTKLATELPRNRY	207	
Db	125	AKAALNHMKNNKINGVNLNISVHQQIRPRVNVQSTSGGKMMETLALNADKGRIV	184	
QY	208	NCLAPGLIKTFSFRLMMDKEEEMKTLR---TRLGEPEDCAGIYSLFSCEDASYTT	264	

Db 185 NALAPGATIAATSN----VDTFKESRSRKOLKKIPKAFGRPEVYAAAAAALVSEASYT 240

QY 265 GETVVGGGT--PSRL 278
| | : | | | : | : |
Db 241 GATLEFVDGGMTLPSQL 257

Search completed: May 8, 2002, 08:27:33
Job time: 192 sec


```

Db 241 RIGEPEDCAGIVSFLCSEDA5YITGETVYVGGTSPSRL 278
|||||
RESULT 2
Q9NV08 PRELIMINARY: PRT: 278 AA.
AC Q9NV08:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE CDNA FLJ11008 FIS. CLONE PLACE1003100, MODERATELY SIMILAR TO HEP27
PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,
RT "NDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
FAMILY.
DR EMBL: AK001870; BAA91953.1;
DR InterPro: IPR002198; Adh_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_1.
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 278 AA; 29509 MM; 6B0096BBA89152A0 CRC64;

Query Match
Best Local Similarity 99.7%; Score 1395; DB 4; Length 278;
Matches 277; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHKAGLGLCARANNVYMASSGMRDPLANKVALYASTDGIQFAIARLADGAAHV 60
|||||
Db 1 MHKAGLGLCARANNVYMASSGMRDPLANKVALYASTDGIQFAIARLADGAAHV 60
QY 61 VSSRKQONVDQAVATLOGEGSLVTGTVCHVQKADRELRVATVAVKLGIDILVNAVN 120
|||||
Db 61 VSSRKQONVDQAVATLOGEGSLVTGTVCHVQKADRELRVATVAVKLGIDILVNAVN 120
QY 121 PFGSINDVEEVDKTLIDINAKAPALMTKAVVPEMEKRGGSVIVSSIAAFSPGFS 180
|||||
Db 121 PFGSINDVEEVDKTLIDINAKAPALMTKAVVPEMEKRGGSVIVSSIAAFSPGFS 180
QY 181 PYNSKATLALGTLTLEIAPRIRVNCIAPGLIKTSFSRMLMDKEESKKEKTLIR 240
|||||
Db 181 PYNSKATLALGTLTLEIAPRIRVNCIAPGLIKTSFSRMLMDKEESKKEKTLIR 240
QY 241 RIGEPEDCAGIVSFLCSEDA5YITGETVYVGGTSPSRL 278
|||||
Db 241 RIGEPEDCAGIVSFLCSEDA5YITGETVYVGGTSPSRL 278

RESULT 3
Q9H3N5 PRELIMINARY: PRT: 260 AA.
AC Q9H3N5:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE NADPH-DEPENDENT RETINOL DEHYDROGENASE/REDUCTASE.

```

```

GN HUMNRD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Furukawa A., Onishi T., Huang D., Araki N., Ichikawa Y.;
RT "CDNA cloning and characterization of peroxisomal short-chain
RT dehydrogenase / reductase that reduce all-trans retinal to retinol.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
FAMILY.
DR EMBL: AB045131; BAB18775.1;
DR InterPro: IPR002198; Adh_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_1.
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 260 AA; 27572 MM; 0BD2E0C0D2E37D08 CRC64;

Query Match
Best Local Similarity 93.0%; Score 1301; DB 4; Length 260;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 MASSGMRDPLANKVALYASTDGIQFAIARLADGAAHVYVSSRKQONVDQAVATLOG 78
|||||
Db 1 MASSGMRDPLANKVALYASTDGIQFAIARLADGAAHVYVSSRKQONVDQAVATLOG 78
QY 79 EGLSTGTGTVCHVQKADRELRVATVAVKLGIDILVNAVNPFSGSINDVEEVDKTL 138
|||||
Db 61 EGLSTGTGTVCHVQKADRELRVATVAVKLGIDILVNAVNPFSGSINDVEEVDKTL 120
QY 139 DINAKAPALMTKAVVPEMEKRGGSVIVSSIAAFSPGFSPIVNSKATLALGTLT 198
|||||
Db 121 DINAKAPALMTKAVVPEMEKRGGSVIVSSIAAFSPGFSPIVNSKATLALGTLT 180
QY 199 ELAPRINIVNCIAPGLIKTSFSRMLMDKEESKKEKTLIRIRLGEPEDCAGIVSFLCSE 258
|||||
Db 181 ELAPRINIVNCIAPGLIKTSFSRMLMDKEESKKEKTLIRIRLGEPEDCAGIVSFLCSE 240
QY 259 DASVITGETVYVGGTSPSRL 278
|||||
Db 241 DASVITGETVYVGGTSPSRL 260

RESULT 4
Q95162 PRELIMINARY: PRT: 260 AA.
AC Q95162:
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE PEROXISOMAL SHORT-CHAIN ALCOHOL DEHYDROGENASE.
GN SCAD-SRL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fransen M., Van Veldhoven P.J., Subramani S.;
RT "Identification of two novel mammalian proteins with a C-terminal
RT peroxisomal targeting signal by using the pvi-phrase display
RT technology.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
FAMILY.
DR EMBL: AF044127; AAD02292.1;

```

DR HSSP: P50163; 2AE1.
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
 DR Oxidoreductase.
 SQ SEQUENCE 260 AA; 27602 MW; 83687CACD82B9BC5 CRC64;

Query Match 92.4%; Score 1293; DB 4; Length 260;
 Best Local Similarity 99.2%; Pred. No. 1.3e-88;
 Matches 258; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 19 MASSGTRRDPPLANKKALVYASTDGIQGFALARRLAODGAHVYSSRKQONVDAVATLQG 78
 DB 1 MASSGTRRDPPLANKKALVYASTDGIQGFALARRLAODGAHVYSSRKQONVDAVATLQG 60
 QY 79 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 138
 DB 61 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 120
 QY 139 DINVKAPALMTKAVVPEMERKGGSVIVSSIAAFSPSPGSPYNSKTLALGLTKTLAI 198
 DB 121 DINVKAPALMTKAVVPEMERKGGSVIVSSIAAFSPSPGSPYNSKTLALGLTKTLAI 180
 QY 199 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 258
 DB 181 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 240
 QY 259 DASYITGETVVGCGTPSRL 278
 DB 241 DASYITGETVVGCGTPSRL 260

RESULT 5
 Q9GKX2 PRELIMINARY; PRT; 260 AA.
 AC Q9GKX2:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NADPH-DEPENDENT RETINOL DEHYDROGENASE/REDUCTASE.
 GN RABNRDR.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER.
 RA Fukukawa A., Ohnishi T., Huang D., Araki N., Ichikawa Y.;
 RT "cDNA cloning and characterization of peroxisomal short-chain
 RT dehydrogenase / reductase that reduce all-trans retinal to retinol.";
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
 CC FAMILY.
 DR EMBL: AB045133; BAB18777.1; .
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
 DR Oxidoreductase.
 KW SEQUENCE 260 AA; 27429 MW; 5B0585B58911B90C CRC64;

Query Match 80.3%; Score 1123; DB 6; Length 260;
 Best Local Similarity 85.4%; Pred. No. 5.6e-76;
 Matches 222; Conservative 18; Mismatches 20; Indels 0; Gaps 0;

QY 19 MASSGTRRDPPLANKKALVYASTDGIQGFALARRLAODGAHVYSSRKQONVDAVATLQG 78
 DB 1 MASSGTRRDPPLANKKALVYASTDGIQGFALARRLAODGAHVYSSRKQONVDAVATLQG 60
 QY 79 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 138
 DB 61 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 120
 QY 139 DINVKAPALMTKAVVPEMERKGGSVIVSSIAAFSPSPGSPYNSKTLALGLTKTLAI 198
 DB 121 DINVKAPALMTKAVVPEMERKGGSVIVSSIAAFSPSPGSPYNSKTLALGLTKTLAI 180
 QY 199 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 258
 DB 181 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 240
 QY 259 DASYITGETVVGCGTPSRL 278
 DB 241 DASYITGETVVGCGTPSRL 260

RESULT 6
 Q9E0U4 PRELIMINARY; PRT; 260 AA.
 AC Q9E0U4:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NADPH-DEPENDENT RETINOL DEHYDROGENASE/REDUCTASE.
 GN MOUNDR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57 BLACK; TISSUE-LIVER.
 RA Fukukawa A., Ohnishi T., Huang D., Araki N., Ichikawa Y.;
 RT "cDNA cloning and characterization of peroxisomal short-chain
 RT dehydrogenase / reductase that reduce all-trans retinal to retinol.";
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
 CC FAMILY.
 DR EMBL: AB045132; BAB18776.1; .
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR Oxidoreductase.
 KW SEQUENCE 260 AA; 27726 MW; 005C5F19AC539CF6 CRC64;

Query Match 77.6%; Score 1086; DB 11; Length 260;
 Best Local Similarity 81.5%; Pred. No. 3.1e-73;
 Matches 212; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

QY 19 MASSGTRRDPPLANKKALVYASTDGIQGFALARRLAODGAHVYSSRKQONVDAVATLQG 78
 DB 1 MASSGTRRDPPLANKKALVYASTDGIQGFALARRLAODGAHVYSSRKQONVDAVATLQG 60
 QY 79 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 138
 DB 61 EGLSVTGTCHVGKADRELRVATAVLAHGGIDILVSNAAVNPFGSIMDYTEVWDKTL 120
 QY 139 DINVKAPALMTKAVVPEMERKGGSVIVSSIAAFSPSPGSPYNSKTLALGLTKTLAI 198
 DB 121 DINVKAPALMTKAVVPEMERKGGSVIVSSIAAFSPSPGSPYNSKTLALGLTKTLAI 180
 QY 199 ELAPRNIRVNCIAPGLIKTSFSRMLMDKEESMKETLRIRRLGPEDCAGIVSFLCSE 258

```

Db 181 ELAPKRNIRVNCIALPGLIKTRFSSVLMMEKAREDEIKEMQIRLGRKPEDCAGIYSFLCSE 240
QY 259 DASITTEETVYVGGGTPSRL 278
      ||||| ||||| ||||| |||||
Db 241 DASINGETVYVGGGTPSRL 260

RESULT 7
ID 099LB2 PRELIMINARY; PRT; 260 AA.
AC 099LB2;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILAR TO PEROXISOMAL SHORT-CHAIN ALCOHOL DEHYDROGENASE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY TUMOR;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
SQ SEQUENCE 260 AA; 27754 MW; 021733159D2BC3C9 CRC64;

Query Match 77.6%; Score 1086; DB 11; Length 260;
Best Local Similarity 81.5%; Pred. No. 3.1e-73;
Matches 212; Conservative 22; Mismatches 26; Indels 0; Gaps 0;

QY 19 MASSGMRTRDPLANKVALVASTDGIQFAIARRLADGAHVYVSSRKQONVDQAVATLQG 78
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MASSGLTRRNPRLSKVALVASTDGIQFAIARRLADGAHVYVSSRKQONVDRAVATLQG 60

QY 79 EGISVCTGCHVCAKEDRELVAFAVLGGIDILVSNAAVNPFGSIMDYTEVMDKTL 138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 EGISVCTGCHVCAKEDRELVAFAVLGGIDILVSNAAVNPFGSIMDYTEVMDKVL 120

QY 139 DINVKAAPALMTKAVPMEKRGGSVYIVSSIAFSPSPSPSPSKTALLGLTKTLAI 198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 SINVTAFAMIKAVPMEKRGGSVYIVSSIAFSPSPSPSKTALLGLTKNFAA 180

QY 199 ELAPRNIRVNCIALPGLIKTRFSSVLMMEKAREDEIKEMQIRLGRKPEDCAGIYSFLCSE 258
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 ELAPKRNIRVNCIALPGLIKTRFSSVLMMEKAREDEIKEMQIRLGRKPEDCAGIYSFLCSE 240

QY 259 DASITTEETVYVGGGTPSRL 278
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 DASINGETVYVGGGTPSRL 260

RESULT 8
ID 09H2R2 PRELIMINARY; PRT; 280 AA.
AC 09H2R2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SHORT-CHAIN ALCOHOL DEHYDROGENASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Pellegrini S., Censini S., Guidotti S., Covacci A., Gabrielli F.;
RL "Human Hep27 chromosomal gene.";
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO THE SHORT-CHAIN ALCOHOL DEHYDROGENASES/REDUCTASES (SDR)
      FAMILY.

```

```

DR EMBL; AF244132; AAC33703.1; -.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR002347; Adh_short_C2.
DR Pfam; PF00106; adh_short_1.
DR Pfam; PF00678; adh_short_C2; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 280 AA; 29926 MW; FE76ED9CB28AB95C CRC64;

Query Match 58.9%; Score 824.5; DB 4; Length 280;
Best Local Similarity 61.6%; Pred. No. 9.9e-54;
Matches 162; Conservative 47; Mismatches 51; Indels 3; Gaps 2;

QY 10 CARANVSRYMSSGTRDPLANKVALVASTDGIQFAIARRLADGAHVYVSSRKQONV 69
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16 CARL--SVRMSSTGIDRGVLANRVAVTGTSIGFAIARRLADGAHVYVSSRKQONV 73

QY 70 DQAVATLQEGLSVIGVCHVCAKEDRELVAFAVLGGIDILVSNAAVNPFGSIMDV 129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 74 DRAMAKLOGEGLSVAGIVCHVCAKEDRELVAFAVLGGIDILVSNAAVNPFGSIMDV 133

QY 130 TEEVMDKTLINVAKAPALMTKAVPMEKRGGSVYIVSSIAFSPSPSPSKTALL 189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 134 SEQIMDKILSYNVKSPALLLSQLLPYEMNR--RGAVIIVSSIAVNPVVALGVSNTAL 192

QY 190 LGLTRTALIELAPRNIRVNCIALPGLIKTRFSSVLMMEKAREDEIKEMQIRLGRKPEDCA 249
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193 LGLTRTALIELAPKDIRNCVAVPGITKTFKSVKVERGNEISLWKNKEHHQRIQISESDCA 252

QY 250 GIVSFLCEDASYITGETVYVGG 272
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 253 GIVSFLCEDASYITGETVYVGG 275

RESULT 9
ID 09D3M7 PRELIMINARY; PRT; 282 AA.
AC 09D3M7;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 5430405K24RIK PROTEIN.
GN 5430405K24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schirini L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RL "Functional annotation of a full-length mouse cDNA collection.";
      Nature 409:685-690(2001).

```

-1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
 CC EMBL: AK017274; BAB30665.1; -
 DR MGD: MGI:1918662; 5430405KRAK.R.
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR002347; Adh_short_C2.
 DR Pfam: PF00106; adh_short; 1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
 KW Oxidoreductase.
 SO SEQUENCE 282 AA; 30070 MW; BB279D84DE7A4DF3 CRC64;

Query Match 56.3%; Score 787; DB 11; Length 282;
 Best Local Similarity 60.7%; Pred. No. 6.2e-51;
 Matches 159; Conservative 32; Mismatches 61; Indels 10; Gaps 2;

QY 16 SVRAASGNTRRDPLANKVALVASTDGGIGFAIRLAODGAHVYSSRKQONVDQAVAT 75
 DB 21 SARRRSKADENRSLAGKVAVITGSTRIGFAIRLAODGAHVYSSRKQONVDQAVIT 80
 QY 76 LOGELSTYGVCHVCAEDREERLVAATKLGIDILVNAVPEFGSIMDYTEWMD 135
 DB 81 LKEEGLSTYGTWCHVGAEDRHLVTLAKHSGGIDELVCVAGVNPVLSITGASQIWD 140
 QY 136 KTLIDNVAPALMTKAVVPEMERGGGSVYVSTIAASPSGFSFYNVSKTALLGLTPT 195
 DB 141 KILDVNVASPALLSKVLPEMENRRGGSVIVSSGVAVVPKLVGYNTSKTALLGLCKS 200
 QY 196 LAIELAPRIRVNCAPGLIKTSFRLMLMDKE-----ESMKELIRLRGCEPDGCG 250
 DB 201 LAVELAPRGIRVNCVLPGLIKTDFLT-----REKTPMNMPLPDMNKIFGVKRLGEPEECAG 255
 QY 251 IVSFLCSSEDASTYTGEEVYVVG 272
 DB 256 IVSFLCSSEDASTYTGEEVYVVG 277

RESULT 10
 Q9D2U3 PRELIMINARY; PRT; 216 AA.

AC Q9D2U3; 01-JUN-2001 (TREMUREL. 17, Created)
 DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)
 DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
 DE DNA SEGMENT, CHR 14, UNIVERSITY OF CALIFORNIA AT LOS ANGELES 2.
 GN D14UC1A2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=CEREBELLUM;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Itawa M., Nishi K., Kiyosawa H., Kondo S., Yamano K.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kusakawa T., Saito I.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schiraldi L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,
 RA Brunschtein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wushaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,

RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
 CC EMBL: AK018788; BAB31411.1;
 DR MGD: MGI:90169; D14UC1A2.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 KW Oxidoreductase.
 SO SEQUENCE 216 AA; 23178 MW; 88FC5B46D020D3E0 CRC64;

Query Match 51.2%; Score 716; DB 11; Length 216;
 Best Local Similarity 76.2%; Pred. No. 8.1e-46;
 Matches 147; Conservative 15; Mismatches 19; Indels 12; Gaps 2;

QY 1 MHKAG-LLGICARANMSVRASSGNTRRDPLANKVALVASTDGGIGFAIRLAODGAHV 59
 DB 1 MOKAGRLGGMTQAMMSVRASSGNTRRDPLANKVALVASTDGGIGFAIRLAODGAHV 60
 QY 60 VVSSRQONVDQAVATLOGGISTGVCHVCAEDREERLVAATKLGIDILVNAV 119
 DB 61 VVSSRQONVDQAVATLOGGISTGVCHVCAEDREERLVAATKLGIDILVNAV 120
 QY 120 NPEFGSIMDYTEWMDKTLIDNVAPALMTKAVVPEMERGGGSVYVSTIAAS 174
 DB 121 NPEFGSIMDYTEWMDKTLIDNVAPALMTKAVVPEMERGGGSVYVSTIAAS 180
 QY 175 -----PSDEGFSF 181
 DB 181 TCRSFHPTSTPP 193

RESULT 11
 Q9VRJ4 PRELIMINARY; PRT; 317 AA.

AC Q9VRJ4; 01-MAY-2000 (TREMUREL. 13, Created)
 DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)
 DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
 DE CG10672 PROTEIN.
 GN CG10672.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle E.G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos K., Doup L.E., Downes M., Dugan-Rocha S., Dew I., Dietz S.M.,
 RA Dudin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keichum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

Matches	136;	Conservative	50;	Mismatches	79;	Indels	2;	Gaps	1;
---------	------	--------------	-----	------------	-----	--------	----	------	----

RC STRAIN=BRISTOL N2;

SQ SEQUENCE 257 AA; 27617 MW; FBF35EDE9DEBEE1F CRC64;

Matches	126;	Conservative	49;	Mismatches	74;	Indels	0;	Gaps	0;
---------	------	--------------	-----	------------	-----	--------	----	------	----

Db 249 INGMHARI 257

Rhabditiidae; Beloderinae; Georhabininae


```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
CC FAMILY.
DR EMBL: 279696; CAB01974.1; -.
DR EMBL: 281592; CAB01974.1; JOINED.
DR EMBL: 281592; CAB04734.1; -.
DR EMBL: 279696; CAB04734.1; JOINED.
DR HSSP: P50163; 2AEL.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_C2.
DR Pfam: PF00678; adh_short_C2.
DR PRINTS: PR00080; SDRFAMILY.
DR Oxidoreductase.
KM SEQUENCE 260 AA; 27590 MW; E43FD36F5EDBA7F4 CRC64;

Query Match
Best local similarity 43.9%; Score 614; DB 5; Length 260;
Matches 120; Conservative 56; Mismatches 70; Indels 4; Gaps 2;

OY 33 KVALVTASTDGIQFAIARLADGAAHVYSSRRKQNVDAVATLQEGSLVTGTVCHVG 91
DB 11 KAIYTAATKKGIGLAIARLDEGASVYIGSRQKNDVAIEYLNKGLTKVAGIAGHTA 70
OY 92 KAEDRELVATAVKLHGIGIDILVNAANVPFGSIMDVTAEVMDKTLIDINKAPALMTRA 151
DB 71 STDQKILVDFLQKFGKINILNNHGINPARGHILEVSDQYWDKLFENVNVAQFGMTL 130
OY 152 VYPEMEKRGSGSVYVSSIAFSPSPGSPYVSKTALLGLTKLAIELAPRINRVNCA 211
DB 131 VAPHTAKESGGAIIFRNASAYKSPPGIAAYGVTKTTLVGLTRALMGLAKONIRVNGIA 190
OY 212 POLIKTSFRMLM--MDKEESMKETLIRRLGPEPCAGIVSFLCSEDAASYITGETV 268
DB 191 PEVITKMSQVLDGDEDAEKELTIOETALGRGLVPPDCACTVATLADSDSYITGEMT 250
OY 269 VVGCGTPSRL 278
DB 251 ITAGGVQARL 260

RESULT 14
O9S9W2 PRELIMINARY; PRT; 254 AA.
AC O9S9W2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE T1J24.9 PROTEIN (AT4G05530 PROTEIN).
GN T1J24.9 OR AT4G05530.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucotids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA All J., Bauer C., Nguyen C., Duckels G.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Washu;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RL Mayer K.F.X.;
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
CC FAMILY.
DR EMBL: AF147263; AAD48959.1; -.
DR EMBL: AL161503; CAB81095.1; -.
DR HSSP: P29132; 1DFT.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR002347; Adh_short_C2.
DR Pfam: PF00106; adh_short_C1.
DR Pfam: PF00678; adh_short_C2.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
DR Oxidoreductase.
KM SEQUENCE 254 AA; 26765 MW; 54297E4DID095372 CRC64;

Query Match
Best local similarity 42.1%; Score 588.5; DB 10; Length 254;
Matches 126; Conservative 50.6%; Pred. No. 3.1e-36;
Mismatches 85; Indels 3; Gaps 2;

OY 30 LANKVALVASTDGIQFAIARLADGAAHVYSSRRKQNVDAVATLQEGSLVTGCH 89
DB 9 LEQKVAIVTASTGIGFEGITERGLEGASVYSSRRQANVDAVAKTSKIDAGYGVCH 68
OY 90 VGKAEDRELVATAVKLHGIGIDILVNAANVPFGSIMDVTAEVMDKTLIDINKAPALMT 149
DB 69 VSNQQRRLNVEKTVKYGKIDILVCAANPSTDPILSSKEVLDKLEINVKSSILL 128
OY 150 KAVYPEMEKRGSGSVYVSSIAFSPSPGSPYVSKTALLGLTKLAIELAPRINRVNCA 209
DB 129 QDMAPHEK--GSSVIFITSAGFSPQAMAMGVTKTALLGLTKLALAMAP-DIRVNA 185
OY 210 LAPGLIKTSFRMLMMDKEESMKETLIRRLGPEPCAGIVSFLCSEDAASYITGETV 269
DB 186 VAPGFVPTHEASFTITGSEVREGIEKTLINRLGTTGDMAAAFAFLADSDSYITGETV 245
OY 270 VVGCGTPSRL 278
DB 246 VAGGMP SRL 254

RESULT 15
O9DCL1 PRELIMINARY; PRT; 141 AA.
AC O9DCL1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DNA SEGMENT, CHR 14, UNIVERSITY OF CALIFORNIA AT LOS ANGELES 2.
GN D14UCLA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,

```

A

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 08:15:25 ; Search time 12.54 Seconds

(without alignments)
498.877 Million cell updates/sec

Title: US-09-866-034-2

Perfect score: 1399

Sequence: 1 MHKAGLLGLCARANNSVMA.....DASTITGETVYVGGTPTSRLL 278

Scoring table:

BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	348	24.9	246	3	US-09-238-481-2
2	335.5	24.0	313	4	US-09-413-814-9
3	328.5	23.5	315	3	US-08-793-035-9
4	328.5	23.5	315	3	US-08-793-035-10
5	324.5	23.2	333	1	US-08-440-856A-4
6	324	23.2	262	4	US-09-363-189B-6
7	321	22.9	247	1	US-08-241-766-13
8	313	22.4	244	1	US-08-375-962B-13
9	313	22.4	244	2	US-08-562-114B-13
10	313	22.4	244	2	US-08-729-594A-13
11	312.5	22.3	256	1	US-08-594-808B-7
12	306.5	21.9	337	3	US-08-440-856A-3
13	299.5	21.4	303	1	US-09-002-298-1
14	296.5	21.2	257	4	US-09-287-097-2
15	295	21.1	244	1	US-08-762-129-1
16	295	21.1	244	1	US-09-090-567-2
17	295	21.1	244	1	US-07-637-865-2
18	294	21.0	244	1	US-08-762-129-3
19	279	19.9	246	6	5229279-7
20	272.5	19.5	255	4	US-08-815-225-4
21	270	19.3	251	3	US-08-822-322-9
22	270	19.3	251	4	US-09-466-109-9
23	270	19.3	252	4	US-08-822-322-8
24	270	19.3	252	4	US-09-466-109-8
25	266	19.0	244	1	US-08-762-129-4
26	258.5	18.5	283	4	US-09-367-012-1
27	253	18.1	273	6	5512669-4

28	248	17.7	359	1	US-08-440-856A-8	Sequence 8, Appli
29	246.5	17.6	248	4	US-09-385-028-11	Sequence 11, Appli
30	241	17.2	335	3	US-09-002-298-7	Sequence 7, Appli
31	237	16.9	295	3	US-09-002-298-5	Sequence 5, Appli
32	230.5	16.5	292	3	US-09-109-705-2	Sequence 2, Appli
33	226	16.2	335	3	US-09-002-298-6	Sequence 6, Appli
34	224	16.0	263	6	5229279-4	Patent No. 5229279
35	221	15.8	335	3	US-09-109-205-19	Sequence 19, Appli
36	217.5	15.5	231	2	US-08-336-198C-7	Sequence 7, Appli
37	217.5	15.5	295	4	US-09-026-482B-2	Sequence 2, Appli
38	197.5	14.1	327	2	US-08-375-962B-12	Sequence 12, Appli
39	197.5	14.1	327	2	US-08-562-114B-12	Sequence 12, Appli
40	197.5	14.1	327	4	US-08-729-594A-12	Sequence 12, Appli
41	191.5	13.7	261	4	US-08-815-225-2	Sequence 12, Appli
42	191.5	13.7	261	4	US-08-815-225-3	Sequence 2, Appli
43	183.5	13.1	388	4	US-08-980-832-41	Sequence 3, Appli
44	173	12.4	325	3	US-08-581-148C-4	Sequence 41, Appli
45	170.5	12.2	309	3	US-09-109-205-1	Sequence 4, Appli
						Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-238-481-2
Sequence 2, Application US/09238481
Patent No. 6110704
GENERAL INFORMATION:
APPLICANT: Huang, Jianzhong
APPLICANT: McDevitt, Damien
TITLE OF INVENTION: Fabg
FILE REFERENCE: GMI0192
CURRENT APPLICATION NUMBER: US/09/238,481
CURRENT FILING DATE: 1999-01-28
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 246
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-238-481-2

Query Match 24.9%; Score 348; DB 3; Length 246;
Best Local Similarity 32.6%; Pred. No. 3.8e-30;

Matches 79; Conservative 57; Mismatches 102; Indels 4; Gaps 3;

QY	33	KVALVTASTDGTGFAIARLADGAAHVVS-SRKOONVQDAVATLOGELSTGTGCHVG	91
DB	5	KSALVTGASRGISGIAQLAEGVNAVAVNAGSKEKAAYVEIKAKGVDSFAIOANVA	64
QY	92	KAEDERLVATVAKLHGIDIIIVSNAAVPPFGSINDVEEYVDRITLDINAKAPALMTKA	151
DB	65	DADDEKAMKEVVSQFSGSDVAVNNAGITR-DNLIMRMEQMDVIDITNLKGVENCIOK	123
QY	152	VPEMEKGGGSGVSVTAAPSPGSPYVNSTALIGLTKTAIEAPNINVCIA	211
DB	124	ATPQHLRQSGAIIMSSVGAIVNGQANVATAGVIGLTKSAARELASRGITVNAVA	183
QY	212	PGLIKTSFRLMMDKEESMKETLIRLGEPECCAGIVSFCSSEASVTGTFVVG	271
DB	184	PGFIYSMDMTDL-SDELKEQMLTQIPLARFGODIDIANVAFSLASDRAKKTYTGOTIHN	241
QY	272	GG 273	
DB	242	GG 243	

RESULT 2
US-09-413-814-9
Sequence 9, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:

```

; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hoeller, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO: 9
; LENGTH: 313
; TYPE: PRF
; ORGANISM: Sorangium cellulosum
; US-09-413-814-9

```

```

Query Match          24.0%; Score 335.5; DB 4; Length 313;
Best Local Similarity 33.6%; Pred. No. 1.3e-28;
Matches 83; Conservative 50; Mismatches 105; Indels 9; Gaps 4;

```

```

QY 30 LANKVALYASTDGGFAIARLADGAAHVYVSSRKQON-VDOAVATVLOGEGSVGTVC 88
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
DB 50 LAGRVALLVYSSRGIGKATLRLAEGADYAVVYHNSKDAEPTALRILGRTRVYQA 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 89 HVGKADRELRVATANKLHGIDILVNAVNEFFESIMDVTEEVMDKTLIDINVKAPALM 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 110 DTRPNAALFSSVPAQGPIDILVNVG-DEFFKPLAAMTDEKVRNVDLSVHVL 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 TRAVVPEMERGGGSVYI--SSIAFSPSPGSPYVNSKTLILGTLKTLATELARINR 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169 CAAVARMORSGRIINIGLSPTIYIRGAPVNAVSIKATGVLITRLATEEAPHGIL 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 VNCGLAPGLIKTSFRMLMDKEKESMKETLIRLGEPEDCAGIVSFCSEDAVITGE 266
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
DB 229 VNCVSPGLIDNGY-----LPPAQKEMERRVPMGRIGRASEVADANAFLASDRASVSGA 283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 267 TVVGGG 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 284 NIAVAGG 290

```

```

RESULT 3
US-08-793-035-9
; Sequence 9, Application US/08793035
; Patent No. 6011201
; GENERAL INFORMATION:
; APPLICANT: Slabas, Antoni R.
; APPLICANT: White, Andrew
; APPLICANT: Chase, Diane
; APPLICANT: Elborough, Keiran
; APPLICANT: Fentem, Phillip A.
; TITLE OF INVENTION: B-ketoacyl ACP Reductase Genes From
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,035
; FILING DATE: 28-JUL-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9414622.2
; FILING DATE: 20-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB95/01678
; FILING DATE: 17-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBY:132
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713.787.1400
; TELEFAX: 713.787.1440
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-793-035-9

```

```

Query Match          23.5%; Score 328.5; DB 3; Length 315;
Best Local Similarity 33.1%; Pred. No. 7.9e-28;
Matches 81; Conservative 53; Mismatches 100; Indels 11; Gaps 5;

```

```

QY 34 VALVYASTDGGFAIARLADGAAHVYV--SSRKQONVDOAVATVLOGEGSVGTVC 89
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
DB 74 VVVVYAGSRIGKATLRLAEGADYAVVYHNSKDAEPTALRILGRTRVYQA 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 VCKAEDRELRVATANKLHGIDILVNAVNEFFESIMDVTEEVMDKTLIDINVKAPALM 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 131 VSKADVAAEMKTAIDAMGTIDVYVNVAGITR-DTLIRKRSQMDVIDNLTVGLCT 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 KAVVPEMERGGGSVYI--SSIAFSPSPGSPYVNSKTLILGTLKTLATELARINR 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 190 QAAFTKIMMKRKGRITINISVGLIGNIGQANYAAKAGVIGFSKTAAREGASRNINAV 249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 LAPGLIKTSFRMLMDKEKESMKETLIRLGEPEDCAGIVSFC-SEDAVITGETV 268
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
DB 250 VCPGFASDMTAKLGEDMEK--ILGTITLGRIGYEDVAGLVERLALSPAASYITGQAF 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 VVGGG 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 308 TIDGG 312

```

```

RESULT 4
US-08-793-035-10
; Sequence 10, Application US/08793035
; Patent No. 6011201
; GENERAL INFORMATION:
; APPLICANT: Slabas, Antoni R.
; APPLICANT: White, Andrew
; APPLICANT: Chase, Diane
; APPLICANT: Elborough, Keiran
; APPLICANT: Fentem, Phillip A.
; TITLE OF INVENTION: B-ketoacyl ACP Reductase Genes From
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX

```

```

Query Match          23.5%; Score 328.5; DB:3; Length 315;
Best Local Similarity 33.1%; Pred. No. 7.9e-26;
Matches 81; Conservative 53; Mismatches 100; Indels 11; Gaps 5;

QY 34 VALVYASTDYGIFAARLAODGAHVYV-----SSRKQONVDQAVATLDEGLSVTGTVCH 89
   | : | : | : | | | | | : | : | : | : | : | : | : | : | : | : |
Db 74 VVVVYAGSRGIGKATLALSLGKAGCVLVNRYAKSAKEAEVSKQIAYGQAITFGG--D 130
   V V V V Y A G S R G I G K A T L A L S L G K A G C V L V N R Y A K S A K E A E V S K Q I A Y G Q A I T F G G -- D

QY 90 VGKAEDREELVATAVAKLHGIDILVSNAAVNPFEFGSIMDVTEEVWDKTLIDINVKAPALMT 149
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 131 VSKREADVAMMKTKAIDAGTIDVYVNNAGITR-DPLLRMKKSQMDVIDNLMTGVFLCT 189
   V S K R E A D V A M M K T K A I D A G T I D V Y V N N A G I T R - D P L L R M K K S Q M D V I D N L M T G V F L C T

QY 150 KAVVPEMKRGGGSVVIYSSIAAFSPSPGSEFYNVSKTALLGLTTLAIEAPNIRVNC 209
   K A V V P E M K R G G S V V I Y S S I A A F S P S P G S E F Y N V S K T A L L G L T T L A I E A P N I R V N C

Db 190 QAAIKIMKKKKRGRILINIASVVLGILGNIGQANFYAAKAGVIGFSTAAAREGASRNINNV 249
   Q A A I K I M K K K R G R I L I N I A S V V L G I L G N I G Q A N F Y A A K A G V I G F S T A A R E G A S R N I N N V

QY 210 LAPGLIKTSFSEBMLMDKESKMETLIRIRLGPEDCAGVSLC-SEDA SYITGETV 268
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 250 VCPGFIASDMTAKIGEDMEKR--ILGTIPLRGYQPEDVAGIVERLALSPASYSITGQAF 307
   V C P G F I A S D M T A K I G E D M E K R -- I L G T I P L R G Y Q P E D V A G I V E R L A L S P A S Y S I T G Q A F

QY 269 VVGGG 273
   V V G G G

Db 308 TIDGG 312
   T I D G G

RESULT 5
US-08-440-856A-4
: Sequence 4, Application US/08440856A
: Patent No. 5750873
GENERAL INFORMATION:
APPLICANT: DELAPORTA, STEPHEN L.
: TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
: TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE. N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA

```

```
Query Match Similarity      23.2%; Score 324.5; DB 1; Length 333;  
Best Local Similarity     32.9%; Pred No. 2.4e-27;  
Matches    102; Conservative   37; Mismatches 108; Indels   63; Gaps    11;
```

OY 13 AMNSVRMASSGCTRRDPLANKVALVTASTDGIFALARIADGAHVVSRSRKOQNVDOA 72
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 37 AMDSPENGAPTPMPKR--LEGGKVAIVTGARGIGEAIVRFVHGAKAVIA-----DIDDA 89

OY 73 VATLOGEGSLV-----TGV-CHYGKAEDREELVATAVLHNGSIDLVSNAAVNPFPG-- 124
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 90 A----GGLLAALAAPHVGVCRCDSVEDEVERAEVARARYGRDLVCNNAGV---LGRQ 142

OY 125 -----SINDVTEENVDKTDLDINVKAPALTFRKVPEMEKGSGSVIYSIAAFSPDPF 179
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 143 TRAAKSILSFPAGEDRVLRNALGALCOMKMRAIALMTORRKRGSIISVASVAGVLCGLCP 202

OY 180 SPYNASKALLGLRTLTALIELAPNRIRVNCLAPGLIKTSFSRLWN----- 225
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 203 HAYTASKEHAIVGLTFKNACCELGHGIRVNCISPFGVATPMLLNARQHGDASPADAADD 262

OY 226 -----DKKEEEMKETLR-----IRLGEPECCAGIVSLFCESDAYITGETVV 268
:||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 263 IDLLDVNSDDQEVEEMEERVROGLATLKCATLR---PRDIAEAFLIASDDSRYTISHNL 318

OY 269 VVGGS-TPSR 277
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 319 VVDGCVTTSR 328

RESULT 6 US-09-363-189B-6
Sequence 6, Application US/09363189B
Patent No. 6242228

GENERAL INFORMATION:
APPPLICANT: SUGIYAMA, MASAKAZU
APPPLICANT: TONOUCHI, NAOTO
APPPLICANT: SUZUKI, SHUNICHI
APPPLICANT: YOKOZEKI, KENZO

TITLE OF INVENTION: XILOPOL DEHYDROGENASE OF ACETIC ACID BACTERIA AND GENE THEREOF
FILE REFERENCE: 0010-1024-0
CURRENT APPLICATION NUMBER: US/09/363, 189B
CURRENT FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: JP10-216047
PRIOR FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 31.0

SEQ ID NO 6
LENGTH: 262
TYPE: PRT
ORGANISM: Gluconobacter oxydans
US-09-363-189B-6

Query Match 23.2%; Score 324; DB 4; Length 262;
Best Local Similarity 30.1%; Pred. No. 1.8e-27;
Matches 78; Conservative 55; Mismatches 104; Indels 22; Gaps 3;

QY 33 KVALVASTDGTGFAIARRLAQDGAHVVSRRQONDAVATLQEGELSVTGVHVK 92
DB 8 KVLVYAGAGNIGLATALRLAEGTAIALDMNRLEAKASVREKGVARSYCDVTS 67
QY 93 AEPRERLVATVTLHGIDILVSNAAVNPFGSIMDTEVWMDKTDINKAKALMTKAV 152
DB 68 EEAVITVDSVADFCKIDLEFNNAAGYQAFAPVDYPSDDFAVLTINTGAHFHLKAV 127
QY 153 VPEMERGGGSVVIVSSIAFSPSPGSPYVNSKTALLGLTKLATELAPRIRVNCAP 212
DB 128 SRQMITQNGRIYNTASMGVKKPMMMAVYGASKAIALTLRALDLAPYIRVNAISP 187
QY 213 GLIKTFSRLMDKEK-----ESMKETLIRRLGEPEDCAGIVSF 254
DB 188 GYVGPFGF---MW-ERQVELQAKVGSQYFSTDPKVVAQOMIGSVPMRRYGDINEIPGVAF 243
QY 255 LCSEDSAYITGETVYVGGG 273
DB 244 LIGDSSFMTGVNLPYAGG 262

RESULT 7

US-08-241-766-13
Sequence 13, Application US/08241766
Patent No. 5686590
GENERAL INFORMATION:
APPLICANT: JACOBS, W. R.
APPLICANT: COLLINS, D. M.
APPLICANT: BANERJEE, A.
APPLICANT: DELISLE, G. W.
APPLICANT: WILSON, T. M.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
TITLE OF INVENTION: AND TREATING MYCOBACTERIAL INFECTIONS USING AN INHA AGENT
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FORBSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/241,766
FILING DATE: 12-MAY-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 25237-20003.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-241-766-13

Query Match 22.9%; Score 321; DB 1; Length 247;
Best Local Similarity 31.9%; Pred. No. 3.6e-27;
Matches 79; Conservative 53; Mismatches 100; Indels 16; Gaps 5;

QY 27 RDLPLANKVALVTASTDGTGFAIARRLAQDGAHVVSRRQONDAVATLQEGELSVTGT 86
DB 10 KPFVSRSLVLTGNGRIGLAIQRLAADGHKAVYTRG-----SGAPKGLFV 58
QY 87 VCHGKAKEDDERLVATVAKLH-GGIDILVSNAAVNPFGSIMDTEVWMDKTDINKAP 145
DB 59 ECDVTSDAVDR-AFTVEEHQGPVEVLVSNAGLSA-DAFLMRTEKFEKVINANLTGA 116
QY 146 ALMTKAVVPEMERGGGSVVIVSSIAFSPSPGSPYVNSKTALLGLTKLATELAPRIR 205
DB 117 FRVAQRASRSMQRKFRMIFIGSVSGWIGQANVAAASKAGVIGMARSIARELSKAVY 176
QY 206 RVNCLAPGLIKTFSRLMDKEKESMKETLIRRLGEPEDCAGIVSFYCSDDASYTG 265
DB 177 TANYVARGYIDTDMTRAL--DERIQGALQIFRAKRVGTAEVAGVYSLASDASYISG 234
QY 266 ETYVYVGGG 273
DB 235 AVIIPVDGG 242

RESULT 8

US-08-375-962B-13
Sequence 13, Application US/08375962B
Patent No. 5731195
GENERAL INFORMATION:
APPLICANT: SIMON, ANDRAS; HELLMAN, UFF; WERNSTEDT,
APPLICANT: CHRISTER, ERIKSSON, UFF.
TITLE OF INVENTION: Isolated Nucleic Acid Molecule
TITLE OF INVENTION: Which Codes for A 32 kDa Protein Having 11-CIS Retinol
TITLE OF INVENTION: Dehydrogenase Activity, and Which Associates With P63,
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect (ASCII standard)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/375,962B
FILING DATE: 20-January-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/258,418
FILING DATE: 6-October-1994
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: LUD 5372.1 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear


```

1 TELEPHONE: (202) 887-1511
2 TELEFAX: (202) 887-0763
3
4 TELE: 706141
5
6 INFORMATION FOR SEQ ID NO: 3
7
8 SEQUENCE CHARACTERISTICS:
9
10 LENGTH: 337 amino acids
11
12 TYPE: amino acid
13
14 STRANDEDNESS: single
15

```


NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0453 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-9845-4166
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:

QY 30 LANKVALYASTDQIGFAIARLLQDGAHVVVSSRKQOONVOAVATLQG--EGLSVGTV 87
 3 LNRRAIYATGNSSIGLQCAEELRAGAEVYINFTDDEDEHAALEKIGRHSVCARYA 62
 QY 88 CHVCAEDREPLVATPAVYKHLGGIDILVLSNAAVNPFEGSIMVTEEVDKTLIDIVKPAAL 147
 Db 63 ADMSDGECRALLETIA---GGCIIIVNNAIQ--HVSIEFFPGKNNAILIINISAFH 111

[illegible]

```

1 TITLE OF INVENTION: chloritetracycline and tetracycline formation and cosmid
2
3 TITLE OF INVENTION: useful therein
4
5 NUMBER OF SEQUENCES: 1
6
7 CORRESPONDENCE ADDRESS:
8
9 ADDRESSEE: American Cyanamid Company
10
11 STREET: One Cyanamid Plaza
12
13 CITY: Wayne
14
15 STATE: New Jersey
16
17 COUNTRY: USA
18
19 ZIP: 07470
20
21 COMPUTER READABLE FORM:
22
23 MEDIUM TYPE: Floppy disk
24
25 COMPUTER: IBM PC compatible
26
27 OPERATING SYSTEM: PC-DOS/MS-DOS
28
29 SOFTWARE: Patentin Release #1.0, Version #1.25
30
31 CURRENT APPLICATION DATA:
32
33 APPLICATION NUMBER: US/08/474,933
34
35 FILING DATE:
36
37 CLASSIFICATION: 435
38
39 PRIOR APPLICATION DATA:
40
41 APPLICATION NUMBER: US 08/125,468
42
43 FILING DATE: 22-SEP-1993
44
45 ATTORNEY/AGENT INFORMATION:
46
47 NAME: Tsevdos, Estelle J
48
49 REGISTRATION NUMBER: 31,145
50
51 REFERENCE/DOCKET NUMBER: 31,255-02
52
53 TELECOMMUNICATION INFORMATION:
54
55 TELEPHONE: (201)831-3241
56
57 TELEFAX: (201)831-3305
58
59 INFORMATION FOR SEQ ID NO: 1:
60
61 SEQUENCE CHARACTERISTICS:
62
63 LENGTH: 30001 base pairs
64
65 TYPE: nucleic acid
66
67 STRANDEDNESS: single
68
69 TOPOLOGY: linear
70
71 MOLECULE TYPE: DNA (genomic)
72
73 US-08-474-933-1

```

[illegible]

```

99  LeuValaIaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSe 115
      |||||.....||| |||||.....||| |||||.....|||
8169 CTGCTGGGGGGGGGGGACCGCTTGGCGCGCGGCTGCGGTGCA 8120
      |||||.....||| |||||.....||| |||||.....|||
115  fAsnAlaIaValAsnProPhePheGlySerIleMetAspValThrGlu 132
      |||||.....||| |||||.....||| |||||.....|||
8119 CAAGCGCGGAGCGAAGCGG...GGTGGCGTCACTCCACCCCTCCAGCG 8073
      |||||.....||| |||||.....||| |||||.....|||
132 LuValITrpaSplyThrLeuAspIleAsnValLysAlaProAlaLeuMet 148
      |||||.....||| |||||.....||| |||||.....|||
8072 AGCTGTGGCAGACGTGATCGACACCACTCACCGCTTCCCGGCTC 8023
      |||||.....||| |||||.....||| |||||.....|||
149 ThrLysAlaValAlaProGlu.....MetGlyLysArgGlyGlyGlySe 163
      |||||.....||| |||||.....||| |||||.....|||
8022 ACCCGCAGGCTCTCACCGCGCGGCGGTGAGCGCGCGCAGCGGCG 7973
      |||||.....||| |||||.....||| |||||.....|||
163 rValValIleValSerSerIleAlaIaPheSerProSerProGlyPheS 180
      |||||.....||| |||||.....||| |||||.....|||
7972 GATCATCAGGCTCCCTCCACCGCGCGCAGCAGGCTGCTCCCGTGGCG 7923
      |||||.....||| |||||.....||| |||||.....|||
180 erProTyAsnValSerLysThrAlaLeuGlyIleThrLysThrLeu 196
      |||||.....||| |||||.....||| |||||.....|||
7922 CCCCTACTCGGCTCCAGCGCGGTATCGGCTTCCACCAAGGCGCTG 7873
      |||||.....||| |||||.....||| |||||.....|||
197 AlaIleGluLeuAlaProArgAsnIleArgValAsnCysLeuAlaProG 213
      |||||.....||| |||||.....||| |||||.....|||
7872 GCCAAGGAACTGCCCCACACCGCGCAGCAGCGTCCGTGCCCGG 7823
      |||||.....||| |||||.....||| |||||.....|||
213 YLeuIleLysThrSerPheSer.....ArgMetL 223
      |||||.....||| |||||.....||| |||||.....|||
7822 CTACGTCGACACCGCGATGGCGGCTGCCGCGTCCCGCAGCGCGCGA 7773
      |||||.....||| |||||.....||| |||||.....|||
223 euTrpMetAspLysGlyLysGlu.....GluSerMetLysGlyThrLeu 237
      |||||.....||| |||||.....||| |||||.....|||
7772 CTGCGGCGACACCGAGGAGGAGGTGCTCGCGCTTCCAGCAGAGATC 7723
      |||||.....||| |||||.....||| |||||.....|||
238 ArgIleArgArgLeuGlyGlyProGluAspCysAlaGlyIleValSerP 254
      |||||.....||| |||||.....||| |||||.....|||
7722 CGCGTGGCGCGTACTCCACCGCGGAGGAGGTGCGCGCATGTCGACT 7673
      |||||.....||| |||||.....||| |||||.....|||
254 eLeuCysSerGluAspAlaSerTyrlleThrGlyGlyThrValValG 271
      |||||.....||| |||||.....||| |||||.....|||
7672 CTGACACACCGCGCGCGGTGCTCACCGCGCAGGCGATCAACGTCT 7623
      |||||.....||| |||||.....||| |||||.....|||
271 LysGlyGly 273
      |||||.....||| |||||.....||| |||||.....|||
7622 GCGGTGCG 7615

```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-238-481-1

```

seq_documentation_block:
; Sequence 1, Application US/09238481
; Patent No. 6110704
; GENERAL INFORMATION:
; APPLICANT: Huang, Jiansheng
; APPLICANT: McDevitt, Damien
; TITLE OF INVENTION: Fabg
; FILE REFERENCE: GM10192
; CURRENT APPLICATION NUMBER: US/09/238,481
; CURRENT FILING DATE: 1999-01-28
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-238-481-1

```

alignment_scores: Quality: 348.00 Length: 242
 Ratio: 1.989 Gaps: 3
 Percent Similarity: 72.314 Percent Identity: 32.645

alignment_block:
 US-09-866-034-2 x US-09-238-481-1

Align seg 1/1 to: US-09-238-481-1 from: 1 to: 741

```

33  LysValAlaIaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAl 49
      ||| |||||.....||| |||||.....||| |||||.....|||
13  AAGAGTGCTTAGTAACAGGTGATCAAGAGAAATGAGACTATATTC 62
      |||||.....||| |||||.....||| |||||.....|||
49  aArgArgLeuAlaGlnAspGlyAlaHisValValSer...SerArgL 65
      |||||.....||| |||||.....||| |||||.....|||
63  GTTCAATTTAGCAGAGAGATTAATGTAGCAGTAACATATGACAGCA 112
      |||||.....||| |||||.....||| |||||.....|||
65  YSGLNGlnAsnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeu 81
      |||||.....||| |||||.....||| |||||.....|||
113  GCAAGAGAAAGCTGAAGCAAGTACTGACAGAAATCAAAAGCTAAAGTGT 162
      |||||.....||| |||||.....||| |||||.....|||
82  SerValThrGlyThrValLysHisValGlyLysAlaGlnAspArgGly 98
      |||||.....||| |||||.....||| |||||.....|||
163  GACAGTTTGGGATTCAGCAAAATGTCCGATGCTGATGATTAAGC 212
      |||||.....||| |||||.....||| |||||.....|||
98  gLeuValAlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValS 115
      |||||.....||| |||||.....||| |||||.....|||
213  AATGATTAAAGAGTACTAGCCAAATTTGGTTTGTAGATGTCTACTAA 262
      |||||.....||| |||||.....||| |||||.....|||
115  eAsnAlaIaIaValAsnProPhePheGlySerIleMetAspValThrGlu 131
      |||||.....||| |||||.....||| |||||.....|||
263  ATAATGCAAGTATTATCTGCG...CATTAATTATTAAATGCGTATGAAAGA 309
      |||||.....||| |||||.....||| |||||.....|||
132  GluValITrpaSplyThrLeuAspIleAsnValLysAlaProAlaLeuMe 148
      |||||.....||| |||||.....||| |||||.....|||
310  CAAGAGTGGGATGATGTTATTGACACCAACTTAAGGCTATTTAAGT 359
      |||||.....||| |||||.....||| |||||.....|||
148  tThrLysAlaValAlaProGluMetGlyLysArgGlyGlySerValY 165
      |||||.....||| |||||.....||| |||||.....|||
360  TATCCAAAAGACACACCAAAATGTTAGCAACAGTAAAGTGTCTATCA 409
      |||||.....||| |||||.....||| |||||.....|||
165  alIleValSerSerIleAlaIaPheSerProSerProGlyPheSerPro 181
      |||||.....||| |||||.....||| |||||.....|||
410  TCAATTTATCAAGTGTGTTGGAGCAGTAAATGCGGACAGCAAGCAAC 459
      |||||.....||| |||||.....||| |||||.....|||
182  TyrAsnValSerLysThrAlaLeuGlyIleThrLysThrLeuAlaI 198
      |||||.....||| |||||.....||| |||||.....|||
460  TATGTTGCACAAAGAGGAGGTATTGTTTAACTTAATCTGGCGGCG 509
      |||||.....||| |||||.....||| |||||.....|||
198  eGluLeuAlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeu 215
      |||||.....||| |||||.....||| |||||.....|||
510  TGAATTAGCATCTCGTGTATCACTTAATGCAAGTGCACCTGGTTT 559
      |||||.....||| |||||.....||| |||||.....|||
215  LelysThrSerPheSerArgMetLeuTrpMetLysGlyLysGlyGlu 231
      |||||.....||| |||||.....||| |||||.....|||
560  TTGTTCTGTGATATGACAGATGCTTTA...AGTATGAGCTTAAAGAA 603
      |||||.....||| |||||.....||| |||||.....|||
232  SerMetLysGlnThrLeuArgIleArgArgLeuGlyGlyProGluAsp 248
      |||||.....||| |||||.....||| |||||.....|||
604  CAATGTTGACTCAAAATTCGTTAGCAGGTTTGGTCAACACAGATAT 653
      |||||.....||| |||||.....||| |||||.....|||
248  saIaGlyIleValSerPheLysSerGlnAspAlaSerTyrlleThrG 265
      |||||.....||| |||||.....||| |||||.....|||
654  TGGTAATACAGTAGCGTCTTACATCAGCAAGCAAAATATATTAACG 703
      |||||.....||| |||||.....||| |||||.....|||
265  LysLuthrValValAlaGlyGlyGly 273
      |||||.....||| |||||.....||| |||||.....|||
704  GTCAACCAATCATGTAAATGTGGA 729

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-239-052-1

```

seq_documentation_block:
; Sequence 1, Application US/09239052
; Patent No. 6346395
; GENERAL INFORMATION:
; APPLICANT: Holmes, David J.

```

```

; APPLICANT: Zhong, Yixi
; APPLICANT: Debouck, Christine
; APPLICANT: Jaworski, Deborah D.
; APPLICANT: Wang, Min
; APPLICANT: Warren, Richard L.
; APPLICANT: Kosmatka, Anna L.
; APPLICANT: McDevitt, Damien
; APPLICANT: Ingraham, Karen A.
; APPLICANT: Chalker, Allison F.
; APPLICANT: So, Chi Young
; APPLICANT: Wallis, Nicola G.
; APPLICANT: Pearson, Stewart C.
; TITLE OF INVENTION: Fabg
; FILE REFERENCE: GM10191
; CURRENT APPLICATION NUMBER: US/09/239,052
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-239-052-1

```

```

alignment_scores:
  Quality: 340.00      Length: 244
  Ratio: 1.954        Gaps: 5
  Percent Similarity: 71.311  Percent Identity: 32.787

```

```

alignment_block:
  US-09-866-034-2 x US-09-239-052-1 ..

```

```

Align seg 1/1 to: US-09-239-052-1 from: 1 to: 732

```

```

30 LeuAlaAsnLysValAlaLeuValThrAlaSerThrAspGlyLeuGly 46
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
7  CTGACACATTAATAATCTTATTACAGCTTCAGCTGCTGCAATTGCTCT 56
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
46 eAlaIleAlaArgArgLeuAlaGlnAspGlyAlaHisValValSerS 63
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
57 TGGCATTCGCCACACAGTTGCTCAGACAGCCACCATTTGCTTAACA 106
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
63 eArgGlySerGlnAsnValAspGlnAlaValAlaThrLeuGlnGly 79
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
107 GTCGTGGGCAATCTCA...GAAGAATTGCTCGCTGAGTTTCAACTAT 153
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
80 GlyLeuSerValThrGlyThrValCysHisValGlyLysAlaGln 96
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
154 GGTATCAAGTGTGCTCCCATTCAGAGATGATACAGATTTGCAAGACC 203
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
96 gGluArgLeuValAlaThrAlaValLysLeuHisGlyGlyLeuAsp 113
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
204 TAAGCTATGATTCATGATCAGCTATTCAGACAGCTGCTGAGTATG 253
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
113 euValSerSerAlaAlaValAsnProPhePheGlySerIleMetasp 129
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
254 TGGTCAACATGACAGGATTACCCAA...GATACCTTATGCTCAGANG 300
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
130 ThrGlnGlnValAlaThrAspLysThrLeuAspIleAsnValLysAla 146
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
301 ACAGAGACGATTTTGAAAAGTGGCTCAAGTCAATGCTGCTGCTCTT 350
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
146 AleuMetThrLysAlaValAlaProGlnMetGlnLysArgGlyGly 163
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
351 TAAATGACACATCATCTTGAACCCGATGTAAGAGCCAGAGAGGAG 400
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
163 eValValIleValSerSerIleAlaAlaPheSerProSerProGlyPhe 179
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
401 CTATCATTAATATGCTACTGTTGTTGCTTATGGGGAATATGCTCAA 450
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
180 SerProTyAsnValSerLysThrAlaLeuLeuGlyLeuThrLysThr 196
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

451 GCTACTATGCTGCTTAAAGCTGGCTTGATGGCTTTACCAAGTCTGT 500
196 uAlaIleGlnLeuAlaProArgAsnIleArgValAsnCysLeuAlaPro 213
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
501 GGCACGGAGGTCCTGACTGCTGAAATGAGAGTCAATGCTGCTCCAG 550
213 LysLeuIleLysThrSerPheSerArgMetLeuTrpMetAspLysGln 229
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
551 GAATGATGAGCTGATATGACACCTATCTTA...TCAGATTAAGATTAG 597
230 GluGlnSerMetLysGluThrLeuArgIleArgArgLeuGlnGluPro 246
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
598 GAACCTACACTAGCTCAG...ATTCGATGAAGAAATTGGCGAGCGAGA 644
246 uAspCysAlaGlyLysValSerPheLeuGlySerGlnLysAlaSerTy 263
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
645 GCAGGTGCGAGATTGTTGACAGTATTTTAAAGCGCCAGAGAT.....TATC 688
263 LeuThrGlyGluThrValValValAlaGlyGly 273
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
689 TAACTGCTCAAGTATGATTCATTTGATGTTGCTGCTG 720

```

```

seq_name: /cgn2_6/ptodata/2/lna/6b_comb.seq:US-09-103-840A-2

```

```

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
; US-09-103-840A-2

```

```

alignment_scores:
  Quality: 462.50      Length: 253
  Ratio: 2.473        Gaps: 4
  Percent Similarity: 73.913  Percent Identity: 39.130

```

```

alignment_block:
  US-09-866-034-2 x US-09-103-840A-2/rev ..

```

```

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

```

```

21 SerSerGlyMetThrArgArgProLeuAlaAsnLysValAlaLeuVal 37
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3071059GCAAGCAATGACCTCCTGAT...CTGACCGCGCCGACCCCTAAT 3071013
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
37 LThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArgArgLeuAla 54
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3071012TACTGGCGCATGCGAGGATCGGCTTGCATCGCCCAACAGCTGGCAG 3070963
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54 LAspGlyAlaHisValValValSerSerArgLysGlnGlnAsnValAsp 70
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3070962CTGCGCGCGCCCATGCTGCTCACCGCACGACGAGGAGAGCTGCCAGC 3070913
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
71 GlnAlaValAlaThrLeuGlnGlnGlyLeuSerValThrGlyThrVal 87

```


Align seg 1/1 to: US-09-103-840A-1 from: 1 to: 4411529

```

10 CysAlaArgAlaTrp.....AsnSerValAr 18
|||||
2193596TGTGTCAGATTGGCTTTCATGTTGGGGCCATCCGAGAACCCAGCA 2193645
18 gmetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysVala 35
|||||
2193646CGTTGGCGGCGGAGTTGACCATCCAGAC...CTGGCGGCGAGGTCG 2193692
35 lalaValAlaThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 51
|||||
2193693CGATCGTTATGGGGGGCGGCGGCGGGAATCGCTGGCGGTTGCCGCGGA 2193742
52 LeuAlaGlnAspGlyAlaHisValAlaValSerSerArgLysGlnAsn 68
|||||
2193743CTGCGCGAGAGGCTGCCATGTGCTGCGCGGACATCGATGCGTATGC 2193792
68 nValAspGlnAlaValAlaThrLeuGlnGlyGluGlyLeuSerValThrG 85
|||||
2193793CGCGGAT...GCCGCGGCCACCAAAATCGGTGTGGC.....G 2193827
85 LysThrValCysHisValGlyLysAlaGluAspArgGlu.....ArgLeu 99
|||||
2193828CAGCGGCTGCGCGGTTGACGTCACGAGACACACAGATCATCGCCATG 2193877
100 ValAlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAs 116
|||||
2193878GTGAGCGCGCTGTTGCCGCTTGGCGGCGGAGTGGACAGTTGGTGGCCA 2193927
116 nAlaAlaValAlaAsnProPhePheGlySerIleMetAspValThrGluGlu 133
|||||
2193928CGCGCGCTGTC...GTTCATCTGCTGCTGCTCATGACACACCGTCGAGG 2193974
133 alThrAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThr 149
|||||
2193975ACTTCGATGGGATCGGATCGGATCAATCTCCGCGCGCTGGCTGCACC 2194024
150 LysAlaValAlaValProGluMetGluLysArgGlyGlyIleGlySerValAla 166
|||||
2194025AAGCATGCGGCAACCGCGGATGATCGAGCGCGCGGCGGAGCCATTGTCAA 2194074
166 eValSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyr 183
|||||
2194075CCGTGCTGTAGCGGCGGAGTACGCGGTGGCGGCGGCGGCGCATACG 2194124
183 snValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlu 199
|||||
2194125GCATGTGCAAGCGCGGATCATCCAGCTCAGCGCATCACCGCGCGGCA 2194174
200 LeuAlaProArgAsnIleArgValAlaAsnLysLeuAlaProGlyLeuLeu 216
|||||
2194175CTGCGCTGCTCGCGGATCGCTCCCAACAGCTGCTCTCCCGCATTCGTCA 2194224
216 sThrSerPheSerArg.....MetLeuThrMetAspLysGluLysGlu 231
|||||
2194225CACCCGATGACAGACCGCGCATGTTCGACGCGGCGCGCTGGGGG 2194274
231 LysMetLysGluThrLeuArgGlyLeuArg.....ArgLeuGlyGlu 244
|||||
2194275CGGCGGCTGCGCGGATCGATGATGCCGCGCTCGAGGCGCGCATGCGCGCA 2194324
245 ProGluAspCysAlaGlyIleValSerPheLeuGlySerGluAspLase 261
|||||
2194325CCGAGAGATGCGCGGATCGCTGCTGCTGCTGCTGCGAGATGCGTC 2194374
261 rTyrIleThrGlyGluThrValAlaValGlyGlyGlyThrProSerArgL 278
|||||
2194375GATGATCACCGGACACCCAGATCGCCGAGCGGCGGAGATTCGCGCC 2194424
278 eu 278
||

```

2194425TC 2194426

seq_name: /cgn2_6/plodata/2/ine/6b_COMB.seq:US-09-468-738A-28

```

seq_documentation_block:
; Sequence 28, Application US/09468738A
; Patent No. 6312933
; GENERAL INFORMATION:
; APPLICANT: Kimoto, No. 6312933 Hiro
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Mitsuhashi, Kazuya
; TITLE OF INVENTION: NOVEL CARBONYL REDUCTASE, METHOD FOR PRODUCING SAID ENZYME, D
; TITLE OF INVENTION: ENCODING SAID ENZYME, AND METHOD FOR PRODUCING ALCOHOL USING
; TITLE OF INVENTION: ENZYME
; FILE REFERENCE: 06501-050001
; CURRENT APPLICATION NUMBER: US/09468738A
; CURRENT FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: JP 1999-171160
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: JP 1998-363130
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0, reformatted using wordperfect 5.1
; SEQ ID NO: 28
; LENGTH: 786
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-468-738A-28

```

alignment_scores: Quality: 331.50 Length: 250
Ratio: 1.973 Gaps: 6
Percent Similarity: 67.200 Percent Identity: 34.000

alignment_block: US-09-866-034-2 x US-09-468-738A-28

Align seg 1/1 to: US-09-468-738A-28 from: 1 to: 786

```

30 LeuAlaAsnLysValAlaLeuValThrAlaSerThrAspGlyIleGlyP 46
|||||
13 TTTAAAGAAAGTCTGCTGCTATTACAGAGCTGCTTACAGGCTCGGAA 62
46 eAlaIleAlaArgArgLeuAlaGlnAspGlyAlaHisValAlaValSer 62
|||||
63 GCGGATGGCCATTCGCTTGGCAAGAGCAAGCAAAAGTGTATCACT 112
63 .....SerArgLysGln.....GlnAsnValAspGlnAlaValAlaThr 75
|||||
113 ATTATGATTAATAACAAGATCCGACAGGATTAAGAAGGATCATCAG 162
76 LeuGlnGlyGluGlyLeuSerValThrGlyThrValCysHisValGly 92
|||||
163 GCGGCGGTGAAGCTGTGCTGCTCAAGAGATGTCAGC.....AA 203
92 sAlaGluAspArgGluArgLeuValAlaThrAlaValLysLeuHisGly 109
|||||
204 AGAGAAAGATGTAATAAATATCTGCACAAACGCGCAATTAAGGACTGCGCA 253
109 LysIleAspIleLeuValSerAsnAlaAlaVal.....AsnProPhePhe 124
|||||
254 CACTGCATATTATGATTAATAAGCCGCGCTTGAATAATCTGTGCCATCT 303
125 SerIleMetAspValThrGluGluValThrAspLysThrLeuAspLase 141
|||||
304 CACGAATATCCGCTCAAGCAT.....TCGATTAAGTCAATCGCGCA 347
141 nValLysAlaProAlaLeuMetThrLys...AlaValAlaProGluMetG 157
|||||
348 CTTAAGGCTGCTTTTGAAGACCGCTGAGAGGATTAATATTGCTAG 397
157 LysArgGlyGlyGlySerValValIleValSerIleAlaAlaPhe 173

```

```

seq_documentation_block:
  Sequence 2, Application US/08793035
  Patent No. 6011201
  GENERAL INFORMATION:
    APPLICANT: Slabas, Antoni R.
    APPLICANT: White, Andrew
    APPLICANT: Chase, Diane
    APPLICANT: Elborough, Keiran
    APPLICANT: Pentem, Phillip A.
    TITLE OF INVENTION: B-Ketocycl ACP Reductase Genes From
    NUMBER OF SEQUENCES: 10
    CORRESPONDENCE ADDRESS:
    ADDRESSEE: Arnold White & Durkee
    STREET: P.O. Box 4433
    CITY: Houston
    STATE: TX
    COUNTRY: US
    ZIP: 77210-4433
    COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/793,035
    FILING DATE: 28-JUL-1997
    CLASSIFICATION: 800
    PRIOR APPLICATION DATA:
    APPLICATION NUMBER: GB 9414622.2
    FILING DATE: 20-JUL-1994
    PRIOR APPLICATION DATA:
    APPLICATION NUMBER: GB95/01678
    FILING DATE: 17-JUL-1995
    ATTORNEY/AGENT INFORMATION:
    NAME: Kammerer, Patricia A.
    REGISTRATION NUMBER: 29,775
    REFERENCE/DOCKET NUMBER: M0T:132
    TELECOMMUNICATION INFORMATION:
    TELEPHONE: 713.787.1400
    TELEFAX: 713.787.1440
    INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
    LENGTH 1155 base pairs

```

604 AGAAGGCGCTGGAGAAATCTTGGCTTCAGTCCGACGACTAGT 913

262 yrllEThrcIygluThrvAlvalvalGlygly 277


```

GENERAL INFORMATION:
APPLICANT: JACOBS, W. R.
APPLICANT: COLLINS, D. M.
APPLICANT: BANERJEE, A.
APPLICANT: DELISLE, G. W.
APPLICANT: WILSON, T. M.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
TITLE OF INVENTION: AND TREATING MYCOBACTERIAL INFECTIONS USING AN INHA AGENT
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/241,766
FILING DATE: 12-MAY-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 25237-20003.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2232 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 494..1234
FEATURE:
NAME/KEY: CDS
LOCATION: 1256..2062
US-08-241-766-12

alignment_scores:
Quality: 321.00 Length: 248
Ratio: 1.922 Gaps: 5
Percent Similarity: 67.339 Percent Identity: 31.855

alignment_block:
US-09-866-034-2 x US-08-241-766-12 ..

Align seg 1/1 to: US-08-241-766-12 from: 1 to: 2232

27 ArgspProleuAlaAsnValAlaLeuValThrAlaSerThrAspGI 43
:::|||||.....:
521 AAACCCCATTCGTATCCCGTCACTCTGTTAACGGAGAAACCGGG 570
43 yllleglyPheAlaIleAlaArgArGLeuAlaGlnAspGIyAlahIsValy 60
|||||||.....:
571 GATCGGGGCTGGCATCGCACAGCGCTGCTGCCGACGGCCACAGGTGG 620
60 alvalserseArGLysGlnGlnAsnValAspGlnAlaValAlaThrLeu 76
::|||::||
621 CGGTACCCACCGCTGA..... 637

77 GlnGlyGlnGlyLeuSerValThrGlyThValAlcYshIsValGlyLysAI 93
:::|||.....:
638 TCCTGGGCGCCCAAAGGGCGCTGTTTGCGTGCATGTGACCTCACCGCAG 687

```

[illegible]

```
351 CCCTCACCCACCGTGA..... 367
      ::|||::: |||
77 glnlygluGlyLeuSerValThrGlyThrValCysHisValGlyLysAl 93
      ::|||::: |||::: |||::: |||
368 TCCGGAGGCCCAAGGGCGTGTGGCGTCGAATGTGACGTCAACCCAG 417
      ::|||::: |||::: |||::: |||
93 agluspargGluArgLeuValAlaThrAlaValLysLeuHis...GlyG 109
      ::|||::: |||::: |||::: |||
418 CGACGCCGTCGATCGC...GCCCTCACGGCGGTAGAGACACACAGGTC 464
      ::|||::: |||::: |||::: |||
109 lylleasplleuValSerAsnAlaValAsnProPheheGlySer 125
      ::|||::: |||::: |||::: |||
465 CGGTGAGGTGCTGGTGTCAACGCCGCGCTATCCGCG...GACGCATTC 511
      ::|||::: |||::: |||::: |||
126 llemetaspValThrGluGluValTrpAspLysThrLeuAspIleAsnVa 142
      ::|||::: |||::: |||::: |||
512 CTCATCCGATGACCCAGCAAGAAAGTTCGAGAAAGTGTATCAACGCCA 561
      ::|||::: |||::: |||::: |||
142 llysAlaProAlaLeuMetThrLysAlaValAlProGluMetGluLysA 159
      ::|||::: |||::: |||::: |||
562 CACCGGGCGGCTCCGGGTGCTCAACGGCATCGCGCAGCATGACAGCGCA 611
      ::|||::: |||::: |||::: |||
159 rgglyGlyLysSerValAlaIleValSerSerIleAlaAlaPheSerPro 175
      ::|||::: |||::: |||::: |||
612 ACAAAATCGGTCAATGATATTCATAGTTCGGTCTCCGCGACCTGGGGC 661
      ::|||::: |||::: |||::: |||
176 SerProGlyPheSerProTyrAsnValSerLysThrAlaLeuLeuGlyLe 192
      ::|||::: |||::: |||::: |||
662 ATCGGCACACAGCCACACTACACGACCTCCAGGCCGAGGTGATGTGCAT 711
      ::|||::: |||::: |||::: |||
192 uThrLysThrLeuAlaIleGluLeuAlaProArgAsnIleArgValAsn 209
      ::|||::: |||::: |||::: |||
712 GGCCTCGCTCGATCGCCCGCGAGCTGTGAGAGCAAACTGACCGCAATG 761
      ::|||::: |||::: |||::: |||
209 yslauAlaProGlyLeuIleLysThrSerPheSerArgMetLeuTrpMet 225
      ::|||::: |||::: |||::: |||
762 TGGTGGCCCCGGGCTACATCGACACCGATATGACCCGCGCTG..... 805
      ::|||::: |||::: |||::: |||
226 AspLysGluLysGluGluSerMetLysGluThrLeuArgIleArgArgLe 242
      ::|||::: |||::: |||::: |||
806 GATGAGCGGATTCACGACGGGGCGCTGCATATTATCCACGAGAGCGGCT 855
      ::|||::: |||::: |||::: |||
242 uGlyGluProGluAspCysAlaGlyIleValSerPheLeuLysSerGlu 259
      ::|||::: |||::: |||::: |||
856 CGGCACCCCGCGCGAGGTGCGCGGGGTGCTGCTGCGCTCCGAGG 905
      ::|||::: |||::: |||::: |||
259 sPlasSerTyrIlePheGlyGluThrValValAlcGlyGly 273
      ::|||::: |||::: |||::: |||
906 ATGCGAGCTATATCTCGGTGCGGTATCCCGGTGACGCGCGGCG 949
```



```

1 MethislysalaglyleuLeuGlyleuCysAlaArgAlaTrpAsnSerVa 17
   ||||| ||| ||||| ||||| ||||| ||| :|||
39 ATGCACATGGCCAGGCTGCTGAGCTCTGTGCTGCGGACGAGAGTGGT 88
   ||||| ||||| ||||| ||||| ||||| ||| :|||
17 lArgMetaLaserSerglyMetThrArgArgAspProLeuAlaAsnLysV 34
   ||||| ||||| ||||| ||||| ||||| ||| :|||
89 GCGGTTGGCCAGCTCCAGATGACCCCGGAGCCGCTCACAATAAGG 138
   ||||| ||||| ||||| ||||| ||||| ||| :|||
34 alAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
   ||||| ||||| ||||| ||||| ||||| ||| :|||
139 TGGGCTGGTACCGGCTCCACCGCGGATCGGCTTCGCCATCGCCGG 188
   ||||| ||||| ||||| ||||| ||||| ||| :|||
51 ArgLeuAlaGlnAspGlyAlaHisValValaValaSerSeraArgLysGlnI 67
   ||||| ||||| ||||| ||||| ||||| ||| :|||
189 CGTTGGCCACGAGGAGGCGCCAGCTGCTGTCAGCAGCGGAGAGCA 238
   ||||| ||||| ||||| ||||| ||||| ||| :|||
67 nasnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSeraValT 84
   ||||| ||||| ||||| ||||| ||||| ||| :|||
239 GAATGTGACGACGAGGCTGCGCAGCTGCAGAGGGAGGGGCTGAGCGTGA 288
   ||||| ||||| ||||| ||||| ||||| ||| :|||
84 hrGlyThrValCysHisValGlyValGlyAlaGlnAspArgIleArgLeuVal 100
   ||||| ||||| ||||| ||||| ||||| ||| :|||
289 CGGGCTGCTGTGCTGATGCGGAGGAGCGGAGACCGGAGCGGCTGGT 338
   ||||| ||||| ||||| ||||| ||||| ||| :|||
101 AlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSeraAla 117
   ||||| ||||| ||||| ||||| ||||| ||| :|||
339 GCCATGGCTGTGAAGCTTCATGAGAGATGATGATCTTCTCCATGTC 388
   ||||| ||||| ||||| ||||| ||||| ||| :|||
117 alAlaValAsnProPhePheGlySerIleMetAspValThrGlyIleValT 134
   ||||| ||||| ||||| ||||| ||||| ||| :|||
389 TGTGTCAACCCCTTCTTGTGAAGCCTAATGATGTCACCGAGAGGTGT 438
   ||||| ||||| ||||| ||||| ||||| ||| :|||
134 rpsAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
   ||||| ||||| ||||| ||||| ||||| ||| :|||
439 GGGACAGACTGTGACATTAAATGTGAAGGCCCGGAGCCCTGATGACAAAG 488
   ||||| ||||| ||||| ||||| ||||| ||| :|||
151 AlaValAlaProGlnMetGlnLysArgGlyGlySerValIleVala 167
   ||||| ||||| ||||| ||||| ||||| ||| :|||
489 GCAGTGGTCCAGAAATGAGAAACAGAGGCGGCTCAGTGTGATGCT 538
   ||||| ||||| ||||| ||||| ||||| ||| :|||
167 lSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnV 184
   ||||| ||||| ||||| ||||| ||||| ||| :|||
539 GTCTTCATGACAGCTTCAGTCAATCTCTGCTTCAGTTCATCAATG 588
   ||||| ||||| ||||| ||||| ||||| ||| :|||
184 alSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlnLeu 200
   ||||| ||||| ||||| ||||| ||||| ||| :|||
589 TCAGTAAACAGCCCTGCTGCGCTCAACATACCTGCGCATAGAGCTG 638
   ||||| ||||| ||||| ||||| ||||| ||| :|||
201 AlaProArgAsnIleArgValaAsnCysLeuAlaProGlyLeuIleLysTr 217
   ||||| ||||| ||||| ||||| ||||| ||| :|||
639 GCCCAAGAGAACTTAGGGTGAAGTGCCT.GCACCTGGACTTATCAAGAC 687
   ||||| ||||| ||||| ||||| ||||| ||| :|||
217 rSerPheSerArgMetLeuTrpMetAspLysGlnLysGlnLysSerMetU 234
   ||||| ||||| ||||| ||||| ||||| ||| :|||
688 TACCTTCAGCAGAGATGCTGTGATGACAAAGAAAGAAAGAAAGCATGA 737
   ||||| ||||| ||||| ||||| ||||| ||| :|||
234 ysgLysThrLeuArgIleArgArgLeuGlyGlnProLysPysCysAlaGly 250
   ||||| ||||| ||||| ||||| ||||| ||| :|||
738 AAGAAACCTGCGGATAGAGAGGTAGCGAGCCAGAGGATGCTGCTGCG 787
   ||||| ||||| ||||| ||||| ||||| ||| :|||
251 lIleValSerPheLeuCysSergLysAspAlaSerTyrIleThrGlyGln 267
   ||||| ||||| ||||| ||||| ||||| ||| :|||
788 ATCGTGTCTTCTGTGCTGTGAAGATGCCAGTACATCAGTGGGAAAC 837
   ||||| ||||| ||||| ||||| ||||| ||| :|||
267 rValValaValGlyGlyThrProSeraArgLeu 278
   ||||| ||||| ||||| ||||| ||||| ||| :|||
838 AGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 871
   ||||| ||||| ||||| ||||| ||||| ||| :|||
seq_name: gb_hlc:BC000663
seq_documentation_block:
LOCUS BC000663 1298 bp mRNA linear HTC 12-Jul-2001
DEFINITION Homo sapiens, Similar to peroxisomal short-chain alcohol

```

```

dehydrogenase, clone IMAGE:3349226, mRNA.
BC000663
VERSION BC000663.1 GI:12653754
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1298)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ruben Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland.
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Meduro, Q.L., Mastaglio, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stanlidop, S., Thomas, P.J.,
Tlionson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LILNL at: http://image.liln.gov
Series: IRAL plate: 5 Row: d Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4105189
This clone has the following problem: frame shifted.
FEATURES
source
1..1298
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3349226"
/tissue="Colon, adenocarcinoma"
/clone_id="NTH MGC_15"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
BASE COUNT 311 a 328 c 400 g 259 t
ORIGIN
alignment_scores:
< Quality: 1289.00 Length: 278
Ratio: 4.756 Gaps: 0
Percent Similarity: 97.482 Percent Identity: 95.324
alignment_block:
US-09-866-034-2 x BC000663
Align seg 1/1 to: BC000663 from: 1 to: 1298
1 MethislysalaglyleuLeuGlyleuCysAlaArgAlaTrpAsnSerVa 17
   ||||| ||| ||||| ||||| ||||| ||| :|||
48 ATGCACATGGCCAGGCTGCTGAGCTCTGTGCTGCGGACGAGAGTGGT 97
   ||||| ||||| ||||| ||||| ||||| ||| :|||
17 lArgMetaLaserSerglyMetThrArgArgAspProLeuAlaAsnLysV 34
   ||||| ||||| ||||| ||||| ||||| ||| :|||
98 GCGGTTGGCCAGCTCCAGATGACCCCGGAGCCGCTCACAATAAGG 147
   ||||| ||||| ||||| ||||| ||||| ||| :|||
34 alAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
   ||||| ||||| ||||| ||||| ||||| ||| :|||

```

```

148 TGGCCCTGCTACGGCCCTCCACCGGAGTCGGCTTCGCCATCCCGG 197
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnI 67
198 CGTTTGGCCAGGAGGAGGCCACAGTGCTGTCTCAGCAGCCGGAGACGCA 247
67 nasValAspGlnAlaValAlaThrLeuGlnGlyGluGlyLeuSerValT 84
248 GAATGTGACAGCGCGTGGCCACGCTGCAGGGGAGGGGCTCAGCGCTGA 297
84 hrcGlyThrValLysHisValGlyLysAlaGlnLysPargGluVal 100
298 CGGGCAGCTGTGTCATGTGGAGCGGAGGACCGGAGCGGCTGCTGT 347
101 AlaThrAlaValLysLeuHisGlyGlyLeuAspIleLeuValSerAsnI 117
348 GCCATGGCTGTGAAGCTTCATGAGGATTCGATATCTTACTTCCATGTC 397
117 AlaValAsnProPhePheGlySerIleMetAspValThrGluGluValT 134
398 TGCTGTCAACCCCTTCTTGTGAGAGCTAATGATGTCACCGAGAGGTGT 447
134 rPasPlyThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
448 GGGACAGACTGTGACATTAAATGTAAGCCCGCCGCTGATGACAAG 497
151 AlaValValProGluMetGluLysArgGlyGlyLysValValIleVal 167
498 GCGCTGTGTCAGAAATGAGAAACAGAGAGGCGCTCAGTGTGATCGT 547
167 LSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnY 184
548 GTCCTTCATAGCAGCCTTCATGTCATCTCCTGCTTCAGTCTTACAAATG 597
184 aLSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
598 TCAGTAAACAGCCCTTGTGCGCTTCACAAACCTCGCCATAGAGCTG 647
201 AlaProArgAsnIleArgValAsnLysLeuAlaProGlyLeuIleLysThr 217
648 GCGCCAGGACACTTAGGGTGAATCTCT. GCACTGTGACTTATCAAGAC 696
217 rSerPheSerArgMetLeuTrpMetAspLysGluLysGluLysSerMetL 234
697 TAGCTTCACAGAGATCTCTGATGAGCAAGAAAGAGAAAGATGA 746
234 ysgLthrThrLeuArgLysArgLysGlyGluProGluAspCysAlaGly 250
747 AAGAAACCTCGCGATMAAGAGTTAGCGACCCAGAGATTGTCTGTGC 796
251 lIeValSerPheLeuLysSerGluAspAlaSerTyrIleThrGlyGluThr 267
797 ATCGTGTCTTCTCTGCTCTGAAGATGCCACTACATCTACTGGGGAAC 846
267 rValValValGlyGlyGlyThrProSerArgLeu 278
847 AGTGTGTGTGCTGAGAGAACCCCGCTCCGCTC 880
seq_name: gb_est2:BE798992
seq_documentation_block:
LOCUS BE798992 782 bp mRNA linear EST 20-SEP-2000
DEFINITION 601583126F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3937471 5',
RNA sequence.
ACCESSION BE798992
VERSION BE798992.1 GI:10220190
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 782)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.

```

```

TITLE
JOURNAL
COMMENT
CONTACT
Tissue Procurement: DCTD/DP
CDNA library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILN at: image.llnl.gov
plate: L1CM783 row: c column: 08
High quality sequence start: 20
High quality sequence stop: 782.
FEATURES
source
1..782
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3937471"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2:
BclRI; CDNA: made by oligo-dt priming. Directionally
cloned into pECORI/XhoI sites using the following 5'
adaptor: GGCACCTAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies). "
BASE COUNT 172 a 200 c 263 g 147 t
ORIGIN
alignment_scores:
Quality: 1257.00 Length: 251
Ratio: 5.008 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-866-034-2 x BE798992
Align seg 1/1 to: BE798992 from: 1 to: 782
3 LysAlaGlyLeuLeuGlyLeuLysAlaArgAlaTrpAsnSerValArgne 19
30 AAGCGGGGCTGCTAGGCTCTGCTGCCGGGCTTGAATTCGTGGTGGAT 79
19 lAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysValAlaI 36
80 GGCACCTCCGGGATACCCCGCGGAGACCCGCTCCGCAAAATAGTGGCC 129
36 euValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArgArgLeu 52
130 TGCTACAGGCTCCACACGAGGATGCGCTTCCATCCCGCGGCTTTG 179
53 AlaGlnAspGlyAlaHisValValValSerSerArgLysGlnLysAsnVa 69
180 GCCCAGAGCGGGGCCCATGTGCTGCACACCGGAGACACACAGATCT 229
69 lAspGlnAlaValAlaThrLeuGlnGlnGlyGluLysLeuSerValThrGlyT 86
230 GGCACAGGCGGGTGGCCACCTCAGGGGAGGGGCGTGCAGACCTGAGCG 279
86 hrValCysHisValGlyLysAlaGlnAspArgGluArgLeuValAlaTrp 102
280 CCCTGTGCCATGTGGGAGGCGGAGACCGGAGCGGCTGTGGCCAGC 329
103 AlaValLysLeuHisGlyLysIleAspIleLeuValSerAsnAlaAlaVa 119
330 GCGTGAACCTTCATGAGGATTCGATATCTCAATGCTGCTGT 379
119 lAsnProPhePheGlySerIleMetAspValThrGluGluValTrpAspL 136

```

```

|||||
380 CAACCCCTTCTTGAAGCATATGATGTCACCTGAGGAGCTGGGACA 429
136 ysthrlleuasplleasvallysalaalaleumethrlysalalaval 152
430 AGACTGTGCACATTATGTCAGAGGCCCAAGCCCTGATGACAAAGCAGCT 479
153 ValProglumetGluysaraglylglyserValValleValserse 169
480 GTGCCAGAAATGAGAAACAGAGGCGGCTCAGTGCTGTCCTTC 529
169 rlelaalaalapheserProserProglyPheSerProTyAsnValserL 186
530 CATAGCAGCCTTCAGTCATCTCCTGCTTCACTCTTCAATGTCAGTA 579
186 ysthrlaaleuenglyleuThrLysThrleuAlailegluleuAlapro 202
580 AAACAGCCTTGTCTGGCCCTGACCAAGACCTGCGCATAGAGCTGGCCCA 629
203 ArgAsnIleArgValAsnGlyleuAlaProglYleuIleLysThrSerph 219
630 AGAAGACTTGGGTGAAGTCTGACAGCCTGAGACTTATCAAGACTGACTT 679
219 eSerArgMetLeuTrpMetAspLysGluysGluSerMetLysGluT 236
680 CACGAGATCTCTGATGAGCAAGAAAAGAGAAAGCATGAAAGAAA 729
236 hrleuarglleargargleuglygluProgluAspCysalaaglyleval 252
730 CCTGGGATTAAGAGATGAGGAGCCAGAGGATGCTGCTGCGCATGCTG 779
253 Ser 253
780 TCT 782

```

seq.name: gb_estL2:BI334186

seq_documentation_block:

LOCUS BI334186 819 bp mRNA linear EST 30-JUL-2001
DEFINITION 602997649FL NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5139733 5',

ACCESSION BI334186

VERSION BI334186.1 GI:15018843

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: c9apbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/MLN at:

http://image.llnl.gov

Plate: LHM11344 Row: a Column: 14

High quality sequence stop: 817.

Location/Qualifiers

1. 819

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5139733"

/clone_lib="NIH_MGC_12"

/tissue_type="cervical carcinoma cell line"

/lab_host="DH10B"

/note="Organ: cervix; Vector: PCMV-SPORE6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

BASE COUNT 186 a 211 c 265 g 157 t
ORIGIN

alignment_scores:

Quality: 1209.50 Length: 267

Ratio: 4.762 Gaps: 2

Percent Similarity: 95.131 Percent Identity: 93.633

alignment_block:

US-09-866-034-2 x BI334186

Align seg 1/1 to: BI334186 from: 1 to: 819

```

1 MethIstysAlaGlyLeuEnglyLeuCysAlaArgAlaTrpAsnSerVa 17
|||||
23 ATGCACATGSCCAGGCTGCTAGGCTTGTGCTGCTGCGCACGAGATCGGT 72
17 ArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysV 34
73 GCGGATGCCAGCTCCAGATGACCCCGGAGACCCGCTCACAATAAGG 122
34 alaAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
123 TGGCCCTGTAAACGGCTCCACGACGAGATCGGCTTGGCATGCGCCGG 172
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnG 67
173 CGTTTGGCCCAAGGACAGAGGCCACGCTGCTGCTGACGACCGAGAGCA 222
67 nasnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValR 84
223 GAATGTGACACGAGCGGCTGCGCCACGCTGCAAGGGAGAGGCTGAGCTGA 272
84 hrGlyThrValLysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
273 CGGGCACTGTGTGCCATGTGGGAGAGCGGAGACCGGAGCGGCTGCTG 322
101 AlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAsnAl 117
323 GCCATGCTGTGAAGCTTCAAGAGGATGATGATATCTCTTCCATGTC 372
117 alaValAlaAsnProPhePheGlySerIleMetAspValThrGluGluValR 134
373 TGTGTTCACACCTTCTTGTGAGAGCCTAATGATGTCACGAGAGAGTGT 422
134 rPAspLysThrLeuAspLysValLysAlaProAlaLeuMetThrLys 150
423 GGGACAAGACTCTGACATTAATGTGAGGCCGCCAGCCCTGATGACAAAG 472
151 AlaValValProglumetGluysaraglylglyserValValleVal 167
473 GCAAGTGTGCCAGAAATGAGAAACAGAGGCGGCTCAGTGTGATGCT 522
167 lSerSerIleAlaAlaPheSerProserProglyPheSerProTyAsnV 184
523 GTTTCATGACAGCCTTCACTGATCACTCTGCTGCTTCACTTACATG 572
184 aLserLysThrAlaLeuEnglyLeuThrLysThrleuAlaIleglule 200
573 TCAGTAAACAGCCTGCTGCGGCTGACCAAGACCTGCGCATGAGA...GC 619
200 uAlaProAlaArgAsnIleArgValAsnGlyleuAlaProglYleuIleLysT 217
620 TGGCCCAAGGAAACATTAGGTGAAGTCCCTAGGACCTGAGCTTATCAAGA 669
217 hrSerPheSerArgMetLeuTrpMetAspLysGluysGluSerMet 233
670 CTAGCTTCAGCAGATGCTTGTGATGACAGAGAAAAGAGAGATG 719
234 LysGluThrLeuArglleargargleuglygluProgluAspCysAlaG 250

```


Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LNCM523 row: 0 column: 16
 High quality sequence start: 13
 High quality sequence stop: 779
 Location/Qualifiers
 1. 974
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3837927"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pORF7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 211 a 243 c 334 g 186 t
 ORIGIN

alignment_scores: Quality: 1198.50 Length: 270
 Ratio: 4.663 Gaps: 6
 Percent Similarity: 95.185 Percent Identity: 94.074

alignment_block:
 US-09-866-034-2 x BE744486

Align seg 1/1 to: BE744486 from: 1 to: 974

8 GlyLeuCySAlaArGaLaTrPaNsSerValArGMeLaSerSerGlyMe 24
 10 GGCCCTGTGTCGGCGCTTGGAAATTCGGTCGGATGCCAGCTCCGGGAT 59
 24 tTtArGaArSpProLeuAlaSnLysValAlaLeuValThAlaSerT 41
 60 GACCCGCGGACCCGCTCCAAATAGAGTGGCCCTGGTAACGGCTCCA 109
 41 hrAspGlyIleGlyPheAlaIleAlaArGArGLeuAlaGlnAspGlyAla 57
 110 CCGACGGGATCGGCTTCGCCATCGCCCGGCTTTGGCCCGAGCGGGGCC 159
 58 HisValAlaValSerSerArGlySGInGlnAsnValAlaSpGlnAlaVal 74
 160 CATGTGCTGTCTCAGACCGCGAAGCAGCAAGTGTGACACGGCGGTGGC 209
 74 aThrLeuGlnGlyGlyLeuSerValThrGlyThrValCysHisValG 91
 210 CAGCGTCAGAGGGAGGGGCTGAGCGTGAACGGCACCCTGCCATGTGG 259
 91 lYlSAlaGluAspArGArGLeuValAlaThrAlaValLysLeuHis 107
 260 GGAAGCGGAGACCGGAGCGGCTGTGGCCACGGCTGTGAAGCTTCAT 309
 108 GlyGlyIleAspIleLeuValSerAsnAlaAlaValAsnProPhePheG 124
 310 GGAGATATGATATCTAGTCTCCATGCTGTCAACCTTTCTTTGG 359
 124 ySerIleMetAspValThrGluValTrpAspLysThrLeuAspIleA 141
 360 AACCAATAAGATGATCTAGAGAGGTGTGGACAAGACTCTGCACATTA 409
 141 snValLysAlaProAlaLeuMetThrLysAlaValAlaProGluMetGlu 157
 |||||||

410 ATGTGAAGGCCCAAGCCCTGATGACAAAGCATGTGTGCCAGAAATGAG 459
 158 LysArGgLyGlySerValValIleValSerSerIleAlaIlePhe 174
 460 AAAGGAGAGGCGGCTCGATGCTGCTTCCATAGACGCTTCAG 509
 174 rProSerProGlyPheSerProTyraSnaValSerLysThrAlaLeuG 191
 510 TCATATCTCTGCTTCACTTCAATATGCTAATAAAGACCTTGTGCTGG 559
 191 lYleuThrLysThrLeuAlaIleGluLeuAlaProArGAsnIleArG 207
 560 GCTGTACCAACACCTGGCATGA...GCTGGCCCAAGAACATTAAGGT 606
 207 lAsnCysLeuAlaProGlyLeuIleLysThrSerPheSerArgMetLeu 224
 607 GAATCGCTTACCTGACCTGATTAACAAGACTAGCTTGAAGGATGCT 656
 224 rMetAspLysGluLysGluLysSerMetLysGluThrLeuArG 240
 657 GGATGACACAGAAACAGAGAAAGCATGAAGAACCCTGGGATAGA 706
 241 ArgLeuGlyGluProGluAspCysAlaGlyIleValSerPheLeuCys 256
 707 AGGTTAGGCGAG...CAGAGGATGTGCTGCATCGCTGTCTTCTGTC 753
 257 SerGluAspAlaSerTyrlleThrGlyGluThrAlaValAlaGlyG 273
 754 TCTGAGATGAGTACATCACTAGGGA...ACCTGTGTGTGGTGGGG 800
 273 yThr 274
 |||||
 801 AAC 804

seq_name: gb_est2:BE741442

seq_documentation_block:
 LOCUS BE741442 755 bp mRNA linear EST 15-SEP-2000
 DEFINITION 601594426F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3948247 5',
 mRNA sequence.
 ACCESSION BE741442
 VERSION BE741442.1 GI:10155434
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 755)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LNCM811 row: d column: 08
 High quality sequence stop: 752.
 Location/Qualifiers
 1. 755
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3948247"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pORF7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average

FEATURES
 source

Insert size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using 2AP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 170 a 200 c 243 g 142 t

ORIGIN

alignment_scores:

Quality: 1169.00 Length: 234
Ratio: 4.996 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-866-034-2 x BE741442

Align seg 1/1 to: BE741442 from: 1 to: 755

```

7 LeuGlyLeuGlyAlaArgAlaTrpAsnSerValArgMetAlaSerSer1 23
CTAGCGCTCTGTGCCGGGCTGTGAATTCGTGCGGATGCGCAGCTCCGG 52
23 yMethrArgArgAspProLeuAlaAsnLysValAlaLeuValThrAla 40
GATGACCGCGCGGACCGCTCGCAATAAAGGTGCCCTGTGTAACGGCCT 102
40 eThrAspGlyIleGlyPheAlaIleAlaArgArgLeuAlaGlnAspGly 56
CCACGCGAGGATCGGCTTCCGATCGCCCGCGCTTTGGCCACAGACGGG 152
103 CCACGCGAGGATCGGCTTCCGATCGCCCGCGCTTTGGCCACAGACGGG 152
57 AlaHisValValSerSerArgLysGlnGlnAsnValAspGlnAlaVal 73
GCCCCATGCTGCTCAGCAGCGGACGAGCATGTGACACGCGCT 202
153 GCCCCATGCTGCTCAGCAGCGGACGAGCATGTGACACGCGCT 202
73 AlaThrLeuGlnGlyLeuGlyLeuSerValThrGlyThrValCysHisV 90
GGCCGCTGCGAGGGGAGGGGCTGAGCGTGCAGCGGACCGCTGCCATG 252
203 GGCCGCTGCGAGGGGAGGGGCTGAGCGTGCAGCGGACCGCTGCCATG 252
90 aGlyLysAlaGlnAspArgGlnArgLeuValAlaThrAlaValLysLeu 106
TGGGGAAGCGGAGGAGCGGCGCTGCTGCCACCGCTTGAAGCTT 302
253 TGGGGAAGCGGAGGAGCGGCGCTGCTGCCACCGCTTGAAGCTT 302
107 HisGlyGlyIleAspIleLeuValSerAsnAlaValAsnProPhePh 123
CATGAGGATCGATATCTCTGATCTCAATGCTCTGCAACCCCTTCTT 352
303 CATGAGGATCGATATCTCTGATCTCAATGCTCTGCAACCCCTTCTT 352
123 eGlySerIleMetAspValThrGlnGlnValTrpAspLysThrLeuAsp 140
TGGAGGCAATATGATGATCTGAGGAGGCTGGGACAAAGACTCTGGACA 402
353 TGGAGGCAATATGATGATCTGAGGAGGCTGGGACAAAGACTCTGGACA 402
140 LeasValLysAlaProAlaLeuMetThrLysAlaValAlaValProGlu 156
TTATATGTAAGGCCCCACCGCTGATGACAAAGCGAGTGTGCCAAGAAATG 452
403 TTATATGTAAGGCCCCACCGCTGATGACAAAGCGAGTGTGCCAAGAAATG 452
157 GluLysArgGlyGlySerValValIleValSerSerIleAlaAlaPh 173
GAGAAACGAGAGGAGGCTCAGTGTGATCGTCTTCATAGACGCCCTT 502
453 GAGAAACGAGAGGAGGCTCAGTGTGATCGTCTTCATAGACGCCCTT 502
173 eSerProSerProGlyPheSerProTyrAsnValSerLysThrAlaLeu 190
CAGTCCATCTCTGCTGCTCACTTACCAATGCTCAATAAAGAGCCCTTGC 552
503 CAGTCCATCTCTGCTGCTCACTTACCAATGCTCAATAAAGAGCCCTTGC 552
190 eugLysLeuThrLysThrLeuAlaIleGlnLeuAlaProArgAsnLys 206
TGGGCTTACCAAGACCTGCGCATAGAGCTGGCCCAAGAGAACTTATAG 602
553 TGGGCTTACCAAGACCTGCGCATAGAGCTGGCCCAAGAGAACTTATAG 602
207 ValAsnCysLeuAlaProGlyLeuIleLysThrSerPheSerArgMet 223
CTGAACCTCCAGACCGCTGATATCAAGACTTACCTTCCAGCAGATCT 652
603 CTGAACCTCCAGACCGCTGATATCAAGACTTACCTTCCAGCAGATCT 652
223 uTrpMetAspLysGlnLysGlnLysMetLysGlnThrLeuArgIleA 240
CTGATGACAAAGGAAAAAGAGAAAGCATGAAAGAAACCTCGCGGATTA 702
653 CTGATGACAAAGGAAAAAGAGAAAGCATGAAAGAAACCTCGCGGATTA 702

```

240 rg 240
11
703 GA 704

seq_name: gb_est2:B1518701

seq_documentation_block:

LOCUS B1518701 967 bp mRNA linear EST 29-AUG-2001
DEFINITION 603062069p1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5211353 5',
mRNA sequence.
ACCESSION B1518701
VERSION B1518701.1 GI:15343493
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 967)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11530 row: 1 column: 18
High quality sequence stop: 841.

FEATURES

source

Location/Qualifiers
1..967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5211353"
/clone_lib="NIH_MGC_118"
/tissue_type="Leukocyte"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 027. Note:
this is a NIH-MGC Library."

BASE COUNT 203 a 256 c 319 g 188 t 1 others

ORIGIN

alignment_scores:

Quality: 1148.00 Length: 288
Ratio: 4.284 Gaps: 13
Percent Similarity: 93.056 Percent Identity: 90.625

alignment_block:

US-09-866-034-2 x B1518701

Align seg 1/1 to: B1518701 from: 1 to: 967

```

1 MethLysValAlaGlyLeuGlnGlyLeuGlyAlaArgAlaTrpAsnSerVal 17
ATGCAACAGGCGGCGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 76
27 ATGCAACAGGCGGCGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 76
17 lArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysV 34
GCGGATGGCCACCTCCGGGATGACCGCGGAGACCCCGCTCGCAATTAAG 126
77 GCGGATGGCCACCTCCGGGATGACCGCGGAGACCCCGCTCGCAATTAAG 126
34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50

```

```

|||||
127 TGGCCCTGTAAAGCGCTCCACGCGGAGTGGCTTGCATCGCCGG 176
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnI 67
|||||
177 CGTTGGCCAGAGCGGGCCATGTGTCTGCTCAGCACCGGAGGACCA 226
67 nasValAspGlnAlaValAlaThrLeuGlnGlyLysLeuSerValT 84
|||||
227 GAATGTGACAGCGGTGGCCAGCTCCAGCGGAGGCGCTGAGCGTGA 276
84 hrcGlyThrValCysHisValGlyLysAlaGlnAspArgLysVal 100
277 CGGCGACCGCTGCCATGTGGGGAAGCGGAGACCGGAGCGCTGGTG 326
101 AlaThrAlaValLysLeuHisGlyLysAlaProAlaLeuMetThrLys 117
327 GCCACGCGCTGTAAGCTTCATGAGGATCGATATCCAGTCCCAATGC 376
117 AlaValAsnProPhePheGlySerIleMetAspValThrLysValT 134
377 TCGTGTCAACCCCTTTCTTGGACCAATATGATGTCTACGAGAGGTGT 426
134 rpaSPlyThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
427 GGGACGAGACTCTGCACATTATGTGAAGGCCCAAGCCCTGATGACAAAG 476
151 AlaValAlaProGluMetGluLysArgGlyLysSerValValIleVal 167
477 GCACTGTGTGCGCAAAATGAGAAACGAGAGCGGCTCAGTGGTCAAT 526
167 L.SerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsn 183
527 TGTCTCCATAGCAGCCTTCAGTCCATCTCCTGCTTCAGTCCCTTACAAT 576
184 ValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlu 200
577 GTGAGTAAACAGCCTTGTGGGCTGACCAAGACCTGGCCATAGAGCT 626
200 euAlaProArg.AsnIleArgValAsnCysLeuAlaProGlyLeuIle 216
627 TGGCCCAAGGAGACATTAGGGTGAACCTGCTACCTGAGCTTATCAA 676
216 sThrSerPheSerArgMetLeuTyr.MetAsp.Lys...GluLysGluI 231
677 GACTAGCTTACGACGAGATGCTGTGATTTGACCAAGGCCAACACAGCA 726
231 uSerMetLysGluThrLeuArgIleArg.ArgLeuGlyGluProGluAsp 247
727 CAGCATGAAGAAACCTCGCGGATMAACAAGGTTAGCGCGACCAAGAT 776
248 CysAlaGly...IleValSerPheLeuCysSerGlu.AspAla.SerTyr 262
777 TGTGCTGGGATCCGCTGCTTCTTCCCTGCTGCTGAAGAAGATGCCAGCT 826
263 IleThr.GlyGluThr.ValValValGly.GlyGlyThrPro 275
827 ATCACTGGGGAAACAGTGTGTGGTGGTGGAGGACCCG 868
seq_name: gb_est2:BG696620
seq documentation block:
LOCUS BG696620 783 bp mRNA linear EST 07-MAY-2001
DEFINITION 602658918f1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4801876 5',
ACCESSION BG696620
VERSION BG696620
KEYWORDS BG696620.1 GI:13961948
SOURCE EST
ORGANISM human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 783)

```

```

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgrabs@emall.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov
Plate: LLNL0695 row: d column: 05
High quality sequence stop: 781.
FEATURES
source
location/Qualifiers
1..783
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4801876"
/clone_lib="NCI_CGAP_Skn3"
/lab_host="DH10B (r1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT0; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 179 a 198 c 256 g 150 t
ORIGIN
alignment_scores:
Quality: 1147.00 Length: 262
Ratio: 4.552 Gaps: 3
Percent Similarity: 96.183 Percent Identity: 93.130
alignment block:
US-09-866-034-2 x BG696620
Align seq 1/1 to: BG696620 from: 1 to: 783
8 GlyLeuCysAlaArgAlaTyrPheSerValArgMetAlaSerSerGlyMet 24
|||||
3 GGCCTGTGCTGTGGCAGCAAGAGTGGTGGGATGGCCAGCTCCAGAT 52
24 tThrArgArgAspProLeuAlaAsnLysValAlaLeuValThrAlaSerT 41
|||||
53 GACCGCGCGGAGACCGCTCAACAATATAGTGGCTGTAAAGGCTTCCA 102
41 hrasGlyIleGlyPheAlaIleAlaArgArgLeuAlaGlnAspGlyAla 57
|||||
103 CCGACGGGATCGGCTTGCATGCGCCGCGGTTGGCCCAAGAGGCC 152
58 HisValValSerSerArgLysGlnGlnAsnValAspGlnAlaValAl 74
|||||
153 CAGCTGTGTCTACGACCGCGAAGCACAGAAATGTGACCAAGCGGTGGC 202
74 aThrLeuGlnGlyGluLysLeuSerValThrGlyThrValCysHisValG 91
|||||
203 CAGCGTCAAGGGGAGAGGCTGACCGTGAACGCGACATGTGCTCATGG 252
91 LysAlaGlnAspArgGluArgLeuValAlaThrAlaValLysLeuHis 107
253 GGAAGGCGGAGACCGCGGAGCGCTGTGTCATATGGCTGTGAAGCTTCA 302
108 GlyGlyLysAlaLeuValSerAsnAlaValAsnProPhePheGln 124
|||||
303 GGAAGTATCATATCTAGTCCAAATGCGTCAACCCCTTCTTGG 352
124 YSerIleMetAspValThrGluGluValTyrPaspLysThrLeuAspIle 141
|||||
353 AACCTTAATGATGTCAACCGAGAGGTGTGGCAAGACTCTGACATTA 402
141 snValLysAlaProAlaLeuMetThrLysAlaValAlaProGluMetGlu 157
|||||

```

403 ATGTGAGGCCCCAGCCCTGATGACAAAGGCACTGTGCCAGAAATGAG 452

158 LysArgGlyGlySerValIleValIleSerSerIleAlaIlePhe 174
 453 AAMCGAGGAGGGCGCTGAGTGATCGTCTTCATAGCAGCCTTCAG 502

174 rProSerProGlyPheSerProTyrAsnValSerLysThrAlaLeuLeu 191
 503 TCCATCTCCCTGGCTTCAGTCTTACAAATGCAATAAAGACCTTCCTG 552

191 LyeuThrLysThrIleuAlaIleGluLeuAlaProArgAsnIleArgVal 207
 553 GCCTCAACATACCTTGCCATAGAGCTGCCCCAAGAACATTAGGGT 602

208 AsnCysLeuAlaProGlyLeuIleLysThrSerPheSerArgMetLeu 224
 603 AACTGCTT.GCAGCTGACTTATCAGAGCTTACAGCAGATGCTCTG 651

224 p.MetAspLysGlyLys.GluGluSerMetLysGluThrLeuArgIle 240
 652 GATTGCAAGCAAGAAACAGAGCAATGAAAGAAAGCCCTCGCATAG 701

240 gArgLeuGlyGluProGlyLysPheCysAlaGlyIle.ValSerPheLys 256
 702 AAGCTTAGCGGAGCCAGAGATTGCTGCGATCGTCTTCTCTGTC 751

257 SerGluAspAlaSerTyrIleThrGlyGlu 266
 752 TCTGAGATGCCAGTACATCACTGGGCA 781

seq_name: gb_est2:BI335522

seq_documentation_block:
 LOCUS BI335522 724 bp mRNA linear EST 30-JUL-2001
 DEFINITION 602997109P1 NIH_MGC_12 Homo sapiens CDNA clone IMAGE:5139338 5',
 mRNA sequence.
 ACCESSION BI335522
 VERSION BI335522.1 GI:15020179
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 724)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/MLN at:
 http://image.llnl.gov
 Plate: LHAM1343 row: a column: 03
 High quality sequence start: 11
 High quality sequence stop: 722.
 Location/Qualifiers
 source
 1. 724
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5139338"
 /clone_lib="NIH_MGC_12"
 /tissue_type="cervical carcinoma cell line"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NotI:
 Site: 2; SalI: Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."
 BASE COUNT 151 a 196 g 138 t
 ORIGIN

alignment_scores:
 Quality: 1105.00 Length: 230
 Ratio: 4.868 Gaps: 2
 Percent Similarity: 98.696 Percent Identity: 98.261

alignment_block:
 us-09-866-034-2 x BI335522

Align seg 1/1 to: BI335522 from: 1 to: 724

1 MethisLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVal 17
 37 ATGCACAGAGCGGGCGGTACTATGCTCTGCGCCGGGCTTGGAATTCGT 86
 17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLys 34
 87 GCGGATGCGCCAGCTCGGGATGACCCGCGGACCCGCTGCAAAATAAG 136
 34 AlaAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
 137 TGGCCCTGGTACGGCTCCACCGAGGATCGGCTTGCCATCGCCGG 186
 51 ArgLeuAlaGlnAspGlyAlaAlaHisValAlaValSerSerArgLysGln 67
 187 CGTTTGGCCAGAGCGGGCCCATGTGTCGTCAGCAGCCGGAAGAGCA 236
 67 nAsnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValTr 84
 237 GAATGTGACCCAGCGGTGCGCCAGCTGCGAGCGGAGGGCTGAGCGTGA 286
 84 hArgLysThrValCysHisValGlyLysAlaGlnAspArgGluArgLeuVal 100
 287 CGGCGACCGTGTCTCCATGTGCGGAGGAGGAGCGGAGCGGCTG 336
 101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerAsnAl 117
 337 GCCACGCGCTGAAAGCTTCATGAGAGTATCATCTTACATCCCAATGC 386
 117 AlaValAsnProPhePheGlySerIleMetAspValThrGluValTr 134
 387 TGTGTTCACCCCTTCTTGTGAAGCATTAATGATGTCACAGAGAGGTGT 436
 134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
 437 GGGACAGACTGTGACATTAAATGTAAGGCCCGCCGATGACAAAG 486
 151 AlaValAlaProGluMetGlyLysArgGlyGlySerValIleVal 167
 487 GCAGTGTGCGCAGAAATGGAAGAACAGAGAGCGGCTCACTGATGATCGT 536
 167 LserSerIleAlaIleAlaPheSerProSerProGlyPheSerProTyrAsn 184
 537 GTCCTTCATAGCAGCCTTCAGTCCATCTCCTGGCTTACAGCTTACATG 586
 184 aLysLysThrAlaLeuLeuGlyLeuThrLysThrIleAlaIleGluLeu 200
 587 TCAGTAATAACAGCCTTCTGCGCTGACCAAGACCCGCGGCAATAGAGCTG 636
 201 AlaProArgAsn.IleArgVal.AsnCysLeuAlaProGlyLeuIleVal 216
 637 GCCCAAGAGAACCTTATGGGTGACAGCTGACCTGACACTTATCAAG 686
 217 ThrSerPheSerArgMetLeuTrpMetAspLysGlu 228
 687 ACTAGCTTACAGAGATGCTGTGATGACAAAGAA 722

seq_name: gb_est2:BG399750

seq_documentation_block:
 LOCUS BG399750 817 bp mRNA linear EST 12-MAR-2001
 DEFINITION 602441351F1 NIH_MGC_75 Homo sapiens CDNA clone IMAGE:4556940 5',

1 Metha1syk1a1g1y1eu1eu1g1y1eu1Cy1a1a1a1g1a1a1r1p1a1s1e1r1v1 17
20 ATGCACACGAGCGGGGCTGCTAGGACCTCTGTCCCGGGCTTGGAAATTCGT 69
17 1A1r1M1e1t1a1s1e1r1S1e1r1G1y1m1e1t1h1a1r1g1a1r1g1a1p1r1o1e1u1a1a1s1y1s1v1 34
70 GCGGATGCGCAGCTCCGGGAGTGAACCCGCGGAGCCCGCTCCCAATAAAG 119
34 a1a1a1e1u1a1l1h1r1a1s1e1r1h1a1s1p1g1y1l1e1g1y1p1h1a1l1a1a1a1r1g1 50
120 TGGCCCTGTGAACGGGCTCCACGAGCGGATGCGCTTGCCATGCCCCCG 166
51 A1r1g1e1u1a1a1a1n1s1p1e1t1y1a1h1s1y1a1y1a1s1e1r1a1r1g1y1s1g1n1g1 67
170 CGTTGGCCCGAGGACGGGCGCCATGTGGTCGACAGCCGGAAGCGACA 219
67 n1a1n1v1a1s1p1a1l1a1y1a1a1e1t1h1e1u1n1g1y1n1g1y1e1u1s1e1r1v1a1t1 84
220 GAATGTGACACGACGGCGGTGCCACCTGCAAGGGGAGGGGCTGTGAGCGTGA 268

CNA Library Preparation: Life Technologies, Inc.
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10386 row: e column: 16
 High quality sequence stop: 777.

FEATURES

source

1..956

Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4508151"
 /clone_id="NIH_MGC_92"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: testis; Vector: PCMV-SPORT6; Site: 1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 212 a 251 c 324 g 167 t 2 others
 ORIGIN

alignment_scores:

Quality: 1102.50 Length: 254
 Ratio: 4.446 Gaps: 2
 Percent Similarity: 97.638 Percent Identity: 94.802

alignment_block:

us-09-866-034-2 x BG257840 ..

Align seg 1/1 to: BG257840 from: 1 to: 956

```

1 Methislysalaglyleuenglyleucysalaatgalaatrpasnserva 17
|||||
7 ATGCACAAAGCGGGGCTGTAGCCCTCTGTGCCCGGCTTGGAATTCGCT 56
|||||
17 largmetalserserlglymetlthararqaspproleualaasnlvsy 34
|||||
57 :CGGATGGCGCAGCTCCGGATGACCCCGCGGAGACCGCTCGCAATTAAG 106
|||||
34 aalaleuvalthralaserthrpsglylleglyphealaallealaary 50
|||||
107 TGGCCCTGTAGCGGCTCCACGAGCGGATCGGCTTCGCATCGCCGG 156
|||||
51 ArgyleAlaGlnaspGlyAlaHisValValSerSerArLySlncl 67
|||||
157 CGTTTGGCCAGGAGCGGCCCATGTGTCAGACAGCGGAGAGAGAGCA 206
|||||
67 nasnvalaspglnalavalathrlenglnlylgluglyleuservalt 84
|||||
207 GAATGTGGACCAAGCGGTGGCCAGCTGAGGGGGGGGGGCTGAGCGTGA 256
|||||
84 hnglythrvalcysrlsvalgllylsalagluaspargluarqleuval 100
|||||
257 CGGGGACCGCTGTGCCATGTGGGGAAGCGGAGACCGGAGCGGCTGTG 306
|||||
101 AlaThrAlaValLysLeuHiscllylIleasplleleuvalSeranaI 117
|||||
307 GCCACGGCTGTGAAGCTTCATGAGGTATCGATATCTAGTCTCCAAATG 356
|||||
117 aalavalasnphephegllyserlmetaspyaltrpugluvalt 134
|||||
357 TCGTGTCAACCCCTTCTTGGAGCATTAATGTGAAGCCCAAGCCCTGATG 406
|||||
134 rPasplysThrleuasplleasvallysalaproalaleumetThrls 150
|||||
407 GGGACAGAGCTCTGGACATTATGTGAAGCCCAAGCCCTGATGACAAAG 456
|||||
151 AlavalalvalproglumetglulysargllylgllyserValValIleVa 167
|||||
457 GCAGTGTGCTCCAGAAATGAGAAAGAGGAGCGGCTCAGTGTATCT 506
|||||
167 lserSerlleAlaIapheserProserProcllyPheSerProTyAsnv 184
|||||

```

seq_name: gb_est2:BG436827

seq_documentation_block:

LOCUS BG436827 695 bp mRNA linear EST 14-MAR-2001

DEFINITION 60248812F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4621082 5',

ACCESSION BG436827

VERSION BG436827.1 GI:13343333

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1- (bases 1 to 695)

AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.

TITLE Unpublished (1999)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaab5-remail.nih.gov

Tissue Procurement: DCRD/DTP/Gazdar

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI383 row: g column: 03

High quality sequence stop: 695.

FEATURES

source

1..695

Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4621082"
 /clone_id="NIH_MGC_18"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pORF7; Site: 1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."
 BASE COUNT 144 a 183 c 231 g 137 t
 ORIGIN

alignment_scores:

Quality: 1097.00 Length: 225
 Ratio: 4.876 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.556

alignment block:
US-09-866-034-2 x BG436827

Align seg 1/1 to: BG436827 from: 1 to: 695

```

1 Methislyslaglyleuenglyleucysalaargalatrpasnserva 17
2 ATGCACAAAGGGGGGCTGCTAGGCTCTGTGCGCGGCTTGGAATTGCGT 71
17 largmetalaasersearlymetthrargaspptoleualasnlvsy 34
72 GCGGATGCGCAGCTCCGGGATGACCCCGCGGACCCCTCGCAATAGG 121
34 alalaleuvalthralsertthraspilylleglyphealaleaarg 50
122 TGGGCTTGTAACGCGCTCCACCGAGGAGATCGGCTTCGCATCGCCGG 171
51 Argleualaglnaspglyalahisvalaialserarglysgingl 67
172 CGTTTGCCCAAGAGCGGGGCCATGTGCTGTCAGCAGCGGAAAGCAGCA 221
67 nasnvalaspglnalalathrleuglnglylglyleuservalt 84
222 GAATGTGACACAGCGGCTGCGCCACGCTCCAGGGGAGGAGGCTGACCTGA 271
84 hrelgylthrvalcyshtsvalglylsalaglunspargluargleuval 100
272 CGGGCACCGGTGCTCCATGTGGGAGAGCGGAGCAGCGGAGCGCTGTG 321
101 Alathralalalysleuhsiglyglylthasplleleuvalserasna 117
322 GCCACGGCTGTGAAGCTTCATGAGGTATGATATCTAGTCTCCATGTC 371
117 aalavalasnphepthehlyserilemetaspvalthrslugluvalt 134
372 TGCTGCAACCTTCTTGGAAGCATATGATGATCTCAGAGAGGTGT 421
134 rpaasplythrleuasplleasnvallysalaprolaleuvalmetthlys 150
422 GGGACAAAGCTCTGACATTAATGTGAAGGCCCGACCGCTGATGACAAAG 471
151 Alalavalproglumetglusargglyglylservalallea 167
472 GCGGTGTGTCAGAAATGAGAAACGAGGAGGCGCTCAGTGTGATCTGT 521
167 lserselealealapheserproserproglipheserprotyrasny 184
522 GTCTTCATAGACGCTTCAGTTCATCTCTGCTTCAGTCTTACCAATG 571
184 alserlysthralaleuenglyleuthrlysthrleualalegluleu 200
572 TCAGTAACAGAGCTTGTGGGCTGACCAAGCCCTGTGCAATAGAGCTG 621
201 Alaprotarasnlllearyvalasnycysleualaprolglyleuileysth 217
622 GC.CCAAGAACATTAGGCTGAACTCCTAGACACTGACTTATCAAGAC 670
217 rserpheserargmetleutripmet 225
671 TAGCTTACACAGAGATGCTCTGATG 695

```

seq_name: gb_est2.BE563882

seq_documentation_block:

LOCUS BE563882 975 bp mRNA linear EST 15-AUG-2000

DEFINITION 601335069P1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3689176 5',

ACCESSION BE563882

VERSION BE563882.1 GI:9807602

KEYWORDS EST.

SOURCE human.

```

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 975)
JOURNAL NIH-MGC http://mgi.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHC383 row: 1 column: 17
High quality sequence start: 82
High quality sequence stop: 734.
FEATURES
source
1. 975
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3689176"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7. Site: 1: XhoI;
Site: 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCGACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT 199 a 254 c 319 g 203 t
ORIGIN
alignment_scores:
Quality: 1096.00 Length: 266
Ratio: 4.455 Gaps: 6
Percent Similarity: 92.481 Percent Identity: 88.722
alignment_block:
US-09-866-034-2 x BE563882
Align seg 1/1 to: BE563882 from: 1 to: 975
1 Methislyslaglyleuenglyleucysalaargalatrpasnserva 17
2 ATGCACAAAGGGGGGCTGCTAGGCTCTGTGCGCGGCTTGGAATTGCGT 112
63 ATGCGCATGCGCGTGGCTGCTAGGCTCTGTGCGCGGCTTGGAATTGCGT 112
17 largmetalaasersearlymetthrargaspptoleualasnlvsy 34
113 GCGGATGCGCAGCTCCGGGATGACCCCGCGGACCCCTCGCAATAGG 162
34 alalaleuvalthralsertthraspilylleglyphealaleaarg 50
163 TGGGCTTGTAACGCGCTCCACCGAGGAGATCGGCTTCGCATCGCCGG 212
51 Argleualaglnaspglyalahisvalaialserarglysgingl 67
213 CGTTTGCCCAAGAGCGGGGCCATGTGCTGTCAGCAGCGGAAAGCAGCA 262
67 nasnvalaspglnalalathrleuglnglylglyleuservalt 83
263 GAATGTGACACAGCTGCGGATGACCCCTCAGGGGAGGAGGCTGACCGT 312
84 ThrGlyThrValCysHisValGlyLysAlaGluAspArgLysGluVal 100
313 ACGGCGACCGTGTGCTATGTGGGAGGAGCGGAGACCGGAGCGGCTGT 362

```

100 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerAsn 117
 117 AlaValAsnProPhePheGlySerIleMetAspValThrGluVal 133
 413 GTCGTCTCAACCTTCTTGGAGCATATGATGATGCTGAGGAGGTG 462
 134 TPASpLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
 463 TGGGACAAAGACTGTCACATTAATGTGAAGGCCCGCCCTGATGACAAA 512
 150 AlaValAlaProGlu.MetGluLysArgGlyGlySerValIle 166
 513 GGCAGTGTGCCAGAAATGGAGAAAGAGGAGCGGCTGAGTGTATC 562
 167 ValSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAs 183
 563 GTCTCTTCATACAGCTTCATCATCTCCTGCTTCAGCTTACAA 612
 183 nValSerLysThrAlaLeuGlyLeuThrLysThrLeuAlaIleGlu 200
 613 TGTCACTAAACAGCTTGTGTGGCTGACCAAGACCTGGCATAGAG. 661
 200 euAla.ProArgAsnIleArg.ValAsnCysLeuAlaProGlyLeuIle 216
 662 ..CTGGCCCGAGAACATTAAGGGTGAAGCTGACACCTGAGCTATATCA 709
 216 yThrSerPheSerArg...MetLeuThrMetLysAspLysGluGlu 231
 710 AACACTAGCTTCAGACAGGATGCTGTGATTGTGACGCGAAACAGAGAA 759
 232 SerMetLysGluThrLeuArgIleArgArgLeuGlyGluProGluSpcy 248
 760 ACCCTCAAGAACCCCTGCGGATTGAAGAGCTTAGCGGACCGAGAGATTG 809
 248 AlaGlyIleValSerPheLeuGlySerGluAspAlaSer 261
 810 TCGGTGCAATCGTGTCTTCTGTGCTGTAAGATGCAGCT 849
 seq_name: gb_est2:B1160793

seq_documentation_block:
 LOCUS B1160793 806 bp mRNA linear EST 05-JUL-2001
 DEFINITION 602864883F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5018877 5',
 mRNA sequence.
 ACCESSION B1160793
 VERSION B1160793.1 GI:14620794
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 806)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LCM1833 row: e column: 22
 High quality sequence stop: 708.
 Location/Qualifiers
 1..806
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5018877"

/clone lib="NIH_MGC_42"
 /tissue_type="epithelioid carcinoma cell line"
 /lab_host="DH108 (phage-resistant)"
 /note="Organ: pancreas; Vector: pOMB7; Site: 1: XhoI;
 Site: 2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GCCACGAGG). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library. 1"

BASE COUNT 182 a 203 c 267 g 154 t
 ORIGIN
 alignment_scores:
 Quality: 1076.00 Length: 272
 Ratio: 4.321 Gaps: 8
 Percent Similarity: 91.544 Percent Identity: 88.235

alignment_block:
 US-09-866-034-2 x B1160793

Align seg 1/1 to: B1160793 from: 1 to: 806

1 MetHisLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaThrAsnSerVa 17
 5 ATGCACAAAGCGGGGCTGCTAGGCGCTCTGTGCCGCGCTTGAATTCGCT 54
 17 LArgMetAlaSerSerGlyMethThrArgArgAspProLeuAlaAsnLys 34
 55 GCGGATGCGCAGCTCCGGATGACCCCGGACCGCGCTGCAAAATTAAG 104
 34 AlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
 105 TGGCCCTGCTAAGCGGCTCCACCGAGGAGTCCGCTTCCGATCGCCCG 154
 51 ArgLeuAlaGlnAspGlyAlaHisValAlaValSerSerArgLysGln 67
 155 CGTTGGCCCGAGGAGCGGCGCCCATGTGCTGTCAGACGCGGAAAGCA 204
 67 nAsnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
 205 GAATGTGGACAGCGGCTGCGACGCTGCAAGGAGGAGGCTGAGCTGA 254
 84 hArgIleThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
 255 CGGGACCGCTGTCTCCATGTGGGAGCGGAGCGGAGCGGCTGCTG 304
 101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerAsn 117
 305 GCCACGCGCTGTGAAGCTTCATGAGGATGATGATATCTCACTCAATG 354
 117 AlaValAsnProPhePheGlySerIleMetAspValThrGluValT 134
 355 TGTCTGTCAACCTTCTTGGAGCATATGATGATGCTGAGGAGGTCT 404
 134 rPASpLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
 405 GGCACAAAGACTGTCACATTAATGTGAAGGCCCGCCCTGATGACAAA 454
 151 AlaValAlaProGluMetGluLysArgGlyGlySerValIleAla 167
 455 GCAGTGTGCCAGAAATGGAGAAAGAGGAGCGGCTGATGATCGT 504
 167 lSerSer.lleAla.lAlaPheSerProSerProGlyPheSerProTyrAs 183
 505 GTCTCTTCATACAGCTTCATCATCTCCTGCTTCAGCTTACAA 554
 183 nValSerLysThrAlaLeuGlyLeuThrLysThrLeuAlaIleGlu 200
 555 TGTCACTAAACAGCTTGTGTGGCTGACCAAGACCTGGCATAGAGC 604

```

200 euAlaProArgAsnIleArgValAsn.CysLeu.AlaProGlyLeuIleL 216
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
605 TGGC.CCAAGGACATTAAGGTGAACATGCTTAGACACCTGACCTTATCA 653
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 ySthrSerPheSerArgMetLeuTrpMetAspLys...GluLysGluGlu 231
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
654 AGATAGCATTCAGCAGGATGCTCTGATGGGACCAAGGAAACCGAGGAA 703
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 SerMet.LysGluThrLeuArgIleArgArgLeuGlyGluProGluAspC 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
704 AGCATGCAAGACACCTGGGATAAGAAAGTTAGCGGACCGACAGAGAA 753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 ysAla...GlyIleValSerPheLeuCysSerGluAspAlaSerTyrIle 263
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
754 GTGTGAATCGGTCTTCTGCTGTCACG.....ATGCAGTATATC 797
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 ThrGly 265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
798 ATGGGG 803

```

About: Results were produced by the Gencore software, version 4.5
Copyright (c) 1993-2000 compugen Ltd.

```
-MODEL=trame+.p2n.model -DEV=xip
-Q/cg12.1/USPTO.spool/US09866034/runtat.08052002.081343_16433/app-query.fasta.1.338
-DB=GenEmul -OEMT=fastap -SUFFFIX=2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cd1 -LIST=45 -DOCALLIGN=200 -THR.SCORE=pct
-THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MUTLEN=0 -MAXLEN=200000000
-USER=US09866034.@CG12.1.6580 -NCPU=6 -ICPU=3 -WALT=LOG
-DEV_TIMEBOOT=120 -WARN_TIMEOUT=30 -NO_XLPU=1 -LONG_THREADS=1
```

Query length: 278

```
Database sequences: 1797656
Database length: 187333701
Search time (sec): 1898.670000
```

```

seq_name: gb_pr:BC003019
seq_documentation_block:
  LOCUS      BC003019              1258 bp    mRNA       linear    PRI 12-JUL-2001
  DEFINITION  Homo sapiens, peroxisomal short-chain alcohol dehydrogenase, clone
  ACCESION   MGC:4052 IMAGE:2822884, mRNA, complete cds.
  VERSION    BC003019.1
  KEYWORDS   GI:12804320
  SOURCE     human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE  1 (bases 1 to 1258) :
             Strausberg, R.
             Direct Submission
             Submitted (06-FEB-2001) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
  REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT    Contact: MGC help desk
             Email: gcgaps@remail.nih.gov
             Tissue Procurement: DCTD/DTF
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
             DNA Sequencing by: Genome Sequence Centre,
             BC Cancer Agency, Vancouver, BC, Canada
             info@cgsc.bc.ca
             Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
             Susanna Chan, Readman Chin, Chris Fjell, Egin Garland, Ran Guin,
             Letticia Hsiao, Martin Kizylinski, Beta Kutsche, Oliver Lee, Soo
             Sen Lee, Victor Ling, Carrie Mathewson, Candice Mcleaver, Steven
             Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
             Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
             Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy,
             George Yang, Scott Zuyderduyn, Marco Marra.
  FEATURES   Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
             Series: IRAL Plate: 9 Row: 1 Column: 21
             This clone was selected for full length sequencing because it
             passed the following selection criteria: matched mRNA gi: 7023406.
             Location/Qualifiers
               1..1258
               /organism="Homo sapiens"
               /db_xref="locusID:10901"
               /db_xref="taxon:9606"
               /clone="MGC:4052 IMAGE:2822884"
               /tissue_type="Lung, small cell carcinoma"
               /clone_id="NIH_MGC-7"
               /lab_host="DH10B-R"
               /note="Vector: pOTB7"
               3..839
               /codon_start=1
               /product="peroxisomal short-chain alcohol dehydrogenase"
               /protein_id="AAH03019.1"
               /db_xref="GI:12804321"
               /translation="MHKAGLLGLCARANNSYRMASGCTRRDPLANKVALVTASTDGI
             GFALARIADGAGHVVVSSRKQWQDQVAVLQEGSVTGVHVGKAEDRERLYAT
             AVKHGIGTIDAFSSNAANPEFGSINDVTEEMDKLDINVAAPALMKAAYPEKRG
             GGSVYISSIAAFSPDGFSPYNSKTAGLGLTFLALIEIAPRINRNVCLAPGLIKTS
             FSRRLMDKREKESMKETLRRRLRGEPEDDCAGIYSLCEDASTITGETVYVGGGTPS
             RL"
BASE COUNT      296 a      318 c      390 g      254 t
ORIGIN

```

alignment_scores:

Quality: 1399.00 Length: 278
 Ratio: 5.032 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-866-034-2 x BC003019

Align seg 1/1 to: BC003019 from: 1 to: 1258

```

1 Methislalsalaglyleuenglyleucysalaaargalatrpasnserva 17
  |||||||
3 ATGCACAAAGGGGGGCTCTAGGCTCTGTCCTCCGGCTTGGAATTCGCT 52
  |||||||
17 largmetlaserseglmetthrargargasprouleualasnysv 34
  |||||||
53 GCGGATGGCCAGCTCCGGGATGACCCCGCGGAGCCCTCCCAATAAAG 102
  |||||||
34 alalaleuvalthrAlaserthrAspglylleglyphealialaarg 50
  |||||||
103 TGGCCCTGTAACGGCTCCACCAAGGATCGCTTGCCATGCCCGG 152
  |||||||
51 Argleualaglnaspglyalalalalalalalalalalalalal 67
  |||||||
153 CGTTTGGCCAGAGGGGGCCATGTGCTCAGACACCGAGACAGCA 202
  |||||||
67 nasnvalasplnalalalalalalalalalalalalalalal 84
  |||||||
203 GAATGTGACCAAGGGGCGGACCGCTCAGGAGGGGCTGAGCGTGA 252
  |||||||
84 hrglythvalalcyshisvalalalalalalalalalalalalal 100
  |||||||
253 CGGCGACCGCTGCCATGTGGGGAAGCGGAGACCGGAGCGCTGCTG 302
  |||||||
101 Alathralalalalysleuhsiglylileasplleuvalserasn 117
  |||||||
303 GCCACGGCTGGAAGCTTCATGAGAGGTGATCATCTCAGTCCAAATG 352
  |||||||
117 aalavalasnprophneglserlilemetasvalthrcluglualt 134
  |||||||
353 TGGTGTCAACCTTTCTTGAACATATGATGATCTCAGTGAAGGCTG 402
  |||||||
134 rpaasplsthrleuasprlleasvalalalalalalalalalal 150
  |||||||
403 GGGACAAAGCTGTGACATTAATGAAAGGCCCAAGCTTATGACAAAG 452
  |||||||
151 AlavalalalProglumetglulysargglylglyservalalale 167
  |||||||
453 GCAGTGTGCGCAAGATGAGAAAGAGAGGCGGCTCAGTGTGATGCT 502
  |||||||
167 lserseerlialalalalalalalalalalalalalalalalal 184
  |||||||
503 GTCTTCCATGACCTTCACTCACTCTGCTCAGTCCATCAATG 552
  |||||||
184 alserlsthAlaleuenglyleuThrlysthrleualalaleglue 200
  |||||||
553 TCAGTAAACAGCTGTGCTGAGCAAGACCTGCGCATAGAGCTG 602
  |||||||
201 AlaproargasnilleargvalalasnCysleualaproglyleuile 217
  |||||||
603 GCGCCAAAGACATTAAGGTGAAGTGTGCTCAGCTGATTAACAAG 652
  |||||||
217 rserpheserargmetleutrpmetasplyglulysleuileu 234
  |||||||
653 TAGCTTCAGCAGATGCTCTGTGATGACAAAGAAAGAGCAATGCA 702
  |||||||
234 ysgluthleuargilleargrleuuglyluproglyaspCysalagly 250
  |||||||
703 AAGAAACCTCGGATAGAGGTTAGCGAGCCAGAGATTGCTGCTG 752
  |||||||
251 lilevalserphleucyssergluaspaalaserlyrillethrclugl 267
  |||||||

```

753 ATCGTCTTTCCTGCTGTGAGATGCCAGCTACATCATCGGGGAAC 802
 267 rValValValGlyGlyGlyThrProSerArgLeu 278
 803 AGTGTGTGTGTGTGAGAGAACCCCTCCGCTC 836

seq_name: gb-Pat:AX201380

seq_documentation_block:

LOCUS AX201380 1283 bp DNA linear PAT 30-AUG-2001
 DEFINITION Sequence 59 from Patent WO0153486.
 ACCESSION AX201380
 VERSION AX201380.1 GI:15391185
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 (bases 1 to 1283)
 Ashkenazi, A.J., Goddard, A., Godowski, P.J., Gurney, A.L.,
 Hillan, K.J., Marsters, S.A., Pan, J., Pittl, R.M., Roy, M.A., Smith, V.,
 Stone, D.M., Watanabe, C.K. and Wood, W.I.
 Compositions and methods for the treatment of tumour
 Patent: WO 0153486A 59 26-JUL-2001;
 Genentech, Inc. (US)

FEATURES
 Location/Qualifiers

source
 1..1283
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 292 a 327 c 401 g 263 t
 ORIGIN

alignment_scores:
 Quality: 1399.00 Length: 278
 Ratio: 5.032 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-866-034-2 x AX201380

Align seg 1/1 to: AX201380 from: 1 to: 1283

```

1 Methislalsalaglyleuenglyleucysalaaargalatrpasnserva 17
  |||||||
36 ATGCACAAAGGGGGGCTCTAGGCTCTGTCCTCCGGCTTGGAATTCGCT 85
  |||||||
17 largmetlaserseglmetthrargargasprouleualasnysv 34
  |||||||
86 GCGGATGGCCAGCTCCGGGATGACCCCGGAGACCGCTGCCAATAAAG 135
  |||||||
34 alalaleuvalthrAlaserthrAspglylleglyphealialaarg 50
  |||||||
136 TGGCCCTGTAACGGCTCCACCAAGGATCGGCTTGCCATGCCCGG 185
  |||||||
51 Argleualaglnaspglyalalalalalalalalalalalalal 67
  |||||||
186 CGTTTGGCCAGAGGGGGCCATGTGCTCAGCAGCCGGAAGCA 235
  |||||||
67 nasnvalasplnalalalalalalalalalalalalalalal 84
  |||||||
236 GAATGTGACCAAGCGGCGGACCGCTCAGGAGGGGCTGAGCGTGA 285
  |||||||
84 hrglythvalalcyshisvalalalalalalalalalalalalal 100
  |||||||
286 GCGGACCGCTGTCCATGTGGGGAAGCGGAGACCGGAGCGCTGCTG 335
  |||||||
101 Alathralalalalysleuhsiglylileasplleuvalserasn 117
  |||||||
336 GCCACGGCTGGAAGCTTCATGAGAGGTGATCATCTCAGTCCAAATG 385
  |||||||
117 aalavalasnprophneglserlilemetasvalthrcluglualt 134
  |||||||

```

```

seq.documentation_block:
LOCUS      AB045131                1325 bp      mRNA      linear      PRI 05-DEC-2000
DEFINITION Homo sapiens humnRDR mRNA for NADPH-dependent retinol
            dehydrogenase/reductase, complete cds.
ACCESSION  AB045131
VERSION    AB045131.1  GI:11559411
SOURCE     Homo sapiens liver cDNA to mRNA.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (sites)
            Furukawa,A., Ohnishi,T., Huang,D., Araki,N and Ichikawa,Y.
            cDNA cloning and characterization of peroxisomal short-chain
            dehydrogenase / reductase that reduce all-trans retinol to retinol
            Unpublished
            2 (bases 1 to 1325)
            Furukawa,A., Huang,D., Ohnishi,T. and Ichikawa,Y.
            Direct Submission
            Submitted (23-JUN-2000) Alzo Furukawa, Kagawa Medical University,
            Faculty of Medicine, Department of Biochemistry, Miki-cho, Ikenobe
            1750-1, Kita-gun, Kagawa 761-0793, Japan
            (E-mail:tfurukawa@kms.sc.jp, Tel:81-87-891-2104, Fax:81-87-891-2105)
FEATURES
     source
         1..1325
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="liver"
            134..916
            /gene="humnRDR"
            134..916
            /gene="humnRDR"
            /codon_start=1

```

protein_id="BAB18775.1"
/db_xref="GI:11559412"
/translation="MASSGMRDRDPLANKVALVYASTDGIGFAIARRLAODGAHVVS
SRKQNVDAVATLOGEGSLVTVGVCHVRAEERERLYATAVKLHGIDILVSNAAVN
PFESIMDVIEVMDKTLIDINVKAPALMAVAVEMKRGGSVIVTSITAAEPSPG
ESPYNKRTALIGLTKTLAELAPRNIRVACLPLGLIKFSFSPRLMDRKESMKT
LRIRLEPEDCAVTSFLLSEDAASYITGETVYVGGTSPRL"
polya_site
BASE COUNT 314 a 333 c 407 g 271 t
ORIGIN

alignment_scores:
Quality: 1399.00 Length: 278
Ratio: 5.032 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-866-034-2 x AB045131 ..

Align seg 1/1 to: AB045131 from: 1 to: 1325

```
1 MethiSLysAlaGlyLeuDeuGlyLeuCysAlaArgAlaTrpAsnSerVa 17
|||||
80 ATGCACAAAGCGCGGGCTGCTAGGCCCTCTGTGCCCCGGCTTGAAATTCGT 129
|||||
17 LArgMetaLaserSergIlyMetLtrArgArgAspProLeuAlaAsnLysV 34
|||||
130 GCGGATGGCCAGCTCCGGGATGACCCGCGGGAGCCGGTCGCAATTAAG 179
|||||
34 aLAlaLeuValAlhrLaserThrAspGlyIleGlyPheAlaIleAlaArg 50
|||||
180 TGGCCCTGGTAAAGCGCTCCACCGAGGAGATCGCTTGGCCATCCCGG 229
|||||
51 ArgLeuAlaGlnAspGlyAlaHisValAlaValSerSerArgLysGlnG 67
|||||
230 CGTTTGGCCCAAGAGCGGGCCCATGTGTCGTCAACACCCGAAACAGCA 279
|||||
67 nAsnValAspGlnAlaValAlaThrLeuGlnGlyLysLeuSerValT 84
|||||
280 GAATGTGACCAAGCGGCGTGGCCACCGCTCAAGGGAGGGGCTGAGCGTGA 329
|||||
84 hGlyThrValIcysHisValGlyLysAlaGlnAspArgGlnArgLeuVal 100
|||||
330 CGGGACCGGTGGCCATGTGGGGAGCGGAGACCGGCGACCGGCTGTGTG 379
|||||
101 AlaThrAlaValLysLeuHisGlyLysIleAspIleLeuValSerAsnAl 117
|||||
380 GCCACGCGTGTGAACCTTCATGAGGATTCGATTCATGCTCCAAAGC 429
|||||
117 aAlaValAsnProPhePheGlySerIleMetAspValThrGlnLysValT 134
|||||
430 TGGCTCAACCTTCTTGGAGAGCATATGATGATCTCACTAGAGAGGTGT 479
|||||
134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
|||||
480 GGGACAAAGACTCTGGACATATATGATGAGAGCGCCAGCCCTGATGCAAA 529
|||||
151 AlaValAlaProGlnMetGlnLysArgGlyLysIleSerValAlaIleVa 167
|||||
530 GCAGGGGTCAGAAATGAGAAACGAGAGGCGGCTCAGTGATCGT 579
|||||
167 lSerSerIleAlaAlaPheSerProSerProGlyPheSerProLysAsnV 184
|||||
580 GTCTTCATACAGACCTTCAGTCCATCTCCGTGCTTCACTTCAATG 629
|||||
184 aLserLysThrAlaLeuGlnGlyLeuThrLysThrLeuAlaIleGlnLeu 200
|||||
630 TCAGTAAACAGCCTTGGCTGGCGCTGACCAAGACCGCTGGCCATGAGACTG 679
|||||
201 AlaProAlaGlnIleArgValAlaAsnCysLeuAlaProGlyLeuIleLysTh 217
|||||
```


680 GCCCCAGAACATTAGGGAAGTGCCTAGCACCTGAGCTTATTCAGAC 729
 217 rSerPheSerArgMetLeuTrpMetAspLysGluLysGluLysMetL 234
 730 TAGCTTCAGCAGAGATCTCTGTGATGACAGAAAAGAGAAACCATGA 779
 224 YSGIuThrLeuArgIleArgArgLeuGlyLysProGluAspCysAlaGly 250
 780 AAGAAACCTCGGATAGAGAGTTAGCGAGCCAGCCAGAGATTGCTGGC 829
 251 lIleValSerPheLeuCysSerGluAspAlaSerTrpIleThrGlyLys 267
 830 ATCGTCTTCTCTGTCTGTGATGACAGCTCATCATCTGGGGAAC 879
 267 rValValValGlyGlyGlyThrProSerArgLeu 278
 880 AGTGGTGGTGGTGGAGAGACCCGCTCCGCTC 913

seq_name: gb_pr:AK001870

seq documentation block:

LOCUS AK001870 1258 bp mRNA linear PRI 22-FEB-2000
 DEFINITION Homo sapiens cDNA FLJ11008 fis, clone PLACE1003100, moderately
 similar to HEP27 PROTEIN.
 ACCESSION AK001870
 VERSION AK001870.1 GI:7023406
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens placenta cDNA to mRNA, clone_1lb:PLACE1
 clone:PLACE1003100.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
 Tanase,T., Nomura,Y., Togiyama,S., Komai,F., Hara,R., Takeuchi,K.,
 Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
 Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuno,Y. and Oshima,A.

TITLE

NEDO human cDNA sequencing project

JOURNAL

Unpublished (2000)

REFERENCE

2 (bases 1 to 1258)
 Isogai,T. and Otsuki,T.
 Direct Submission

AUTHORS

Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yata, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

COMMENT

International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction; 5'- & 3'-end one pass sequencing and clone selection;
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES

Location/Qualifiers
 1..1258

CDS

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="PLACE1003100"
 /cissue_type="Placenta"
 /clone_1lb="PLACE1"
 /note="cloning vector: pME18SFL3"
 21..857
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91953.1"
 /db_xref="GI:7023407"

BASE COUNT

282 a 322 c 394 g 260 t

ORIGIN

alignment_scores:
 Quality: 1395.00 Length: 278
 Ratio: 5.018 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.640

alignment_block:
 US-09-866-034-2 x AK001870

Align seg 1/1 to: AK001870 from: 1 to: 1258

1 MethIstLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVa 17
 21 ATGCACAAGCGCGGCTGCTAGCCTCTGTGCTCCCGGCTGGAAATTCGCT 70
 17 lArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysV 34
 71 GCGATGGCCACGCTCCGGATGACCCCGCGGAGCCGCTCGCAATAAGG 120
 34 aAlaIleValThrAlaSerTrpAspGlyIleGlyPheAlaIleAlaArg 50
 121 TGGCCCTGGCAAGCGGCTCCACCGAGCGGATCGCTTGCCATCCCGCG 170
 51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnG 67
 171 GCTTGGCCCAAGAGCGGGCCCATGTGCTGCTACAGCAGCCGAGACACA 220
 67 nasValAspGlnAlaValAlaThrLeuGlnGlyLysLysSerValT 84
 221 GATGTGACACAGCGGCGTGGCCAGCGCTCAGGGGAGGGGCTGACCGTGA 270
 84 hrGlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
 271 CGGCGACCGGTGGCCATGCGGGAAGCGGAGACCGGAGCGGTGGTG 320
 101 AlaThrAlaValLysLeuHisGlyLysLysLysLysLysLysLysLys 117
 321 GCCACGGCTGTGAACCTTCATGAGATGATGATGATGATGATGATGATG 370
 117 aAlaValAsnProPhePheGlySerIleMetAspValThrGluGluValT 134
 371 TGCTTCACACCTTTCTTTGGAGCATATGATGATGATGATGATGATGATG 420
 134 rPaspLysThrLeuAspLysLysValLysAlaProAlaLeuMetThrLys 150
 421 GCGACAAGACTCTGACATTAATGATGAGAGCCCGCCGCTGATGACAAG 470
 151 AlaValValProGluMetGluLysArgGlyLysLysValValIleVal 167
 471 GCAGTGGTGGCCAGAAATGAGAAACGAGAGCGGCTCAGTGGTATCGT 520
 167 lSerSerIleAlaAlaPheSerProSerProGlyLysPheSerProTrpAsnV 184
 521 GTCTTCATACAGCTTCTTCTTGGAGCATATGATGATGATGATGATGATG 570
 184 aLysLysThrAlaLeuLeuGlyLeuThrLysThrLysAlaIleGluLeu 200
 571 TCAGTAAACAGCCTTGTGGGCTGACCAAGACCTTGCCATGAGACTG 620
 201 AlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysTh 217
 621 GCCCAAGACCATTAAGGTGAGTGCCTAGCCTGACCTGACCTTATTCAGAC 670
 217 rSerPheSerArgMetLeuTrpMetAspLysGluLysGluLysMetL 234
 671 TAGCTTCAGCAGAGATCTCTGTGATGACAGAAAAGAGAAACCATGA 720
 234 YSGIuThrLeuArgIleArgArgLeuGlyLysProGluAspCysAlaGly 250
 721 AAGAAACCTCGGATAGAGAGTTAGCGAGCCAGCCAGAGATTGCTGGC 770

```

251 ILevalSerPheLeucysSerGluaspAlaSerTyrIleThrcGlyLurh 267
|||||
771 ATCGTCTCTTCTCTGCTCTGAAGATGCCAGTACATCATCGGGGAAC 820
|||||
267 rValValValGlyGlyThrProSerArgLeu 278
|||||
821 AGTGGTGGTGGTGGAGAACCCCGTCCGCCCTC 854

seq_name: gb_pr:AF044127
seq_documentation_block:
LOCUS AF044127 1281 bp mRNA linear PRI 27-MAY-1999
DEFINITION Homo sapiens peroxisomal short-chain alcohol dehydrogenase
(SCAD-SRL) mRNA, complete cds.
ACCESSION AF044127
VERSION AF044127.1 GI:4105189
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1281)
AUTHORS Franssen, M., Van Veldhoven, P.P. and Subramani, S.
TITLE Identification of peroxisomal proteins by using M13 phage protein
VI phage display: molecular evidence that mammalian peroxisomes
contain a 2,4-dienoyl-CoA reductase
Biochem. J. 340 (Pt 2), 561-568 (1999)
JOURNAL Biochem. J. 340 (Pt 2), 561-568 (1999)
MEDLINE 99267333
REFERENCE 2 (bases 1 to 1281)
AUTHORS Franssen, M. and Subramani, S.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-1998) Biology, University of California, San
Diego, Gilman Drive, La Jolla, CA 92093-0322, USA
FEATURES
Source
1..1281
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1281
/gene="SCAD-SRL"
misc_feature
1..1281
/gene="SCAD-SRL"
/note="similar to ESTs with GenBank Accession Numbers
R25850, AA293652 and H63672"
80..862
/gene="SCAD-SRL"
/note="SCAD-SRLP; contains a C-terminal peroxisomal
targeting signal (PTS1)"
/codon_start=1
/product="peroxisomal short-chain alcohol dehydrogenase"
/db_xref="GI:4105190"
/translaton="MASSGMRDRDPLANKVALVYASTDGIQFAIARLAADGAHVVS
SRQQNVDAVAVTLOEGLSVTGVCHVGAEDREIRLVAAVKLHGIDILVNAAVN
PFGSLMDVTEVWEDKTDINVKAPALMTKAVVPEMKRGGSVIVISIIAASPSPG
FSFYNSKTKALGTLTALTEIAPRINRVNCLAPGLIKTSPSRMLMDKEKESMKET
LRIRLRGEPPDCAGIYSFLCSEASITTGFTVVVGGTSPSRL"
polyA_signal
1235..1240
/gene="SCAD-SRL"
BASE COUNT 300 a 323 c 395 g 263 t
ORIGIN
alignment_scores:
Quality: 1391.00 Length: 278
Ratio: 5.022 Gaps: 0
Percent Similarity: 99.640 Percent Identity: 99.281
alignment_block:
US-09-866-034-2 x AF044127 ..
Align seg 1/1 to: AF044127 from: 1 to: 1281
1 MethiSLySAIaGlyLeuLeuGlyLeuGlySAlaAraGAlaTAPaNSerVa 17

```

```

26 ATGCACAGGCGGGGCTGTAGGCTCTTGCCCGGCTTGGAATTCGGT 75
|||||
17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaSLysV 34
|||||
76 GCGGATGGCCAGCTCCGGATGACCCCGCGGACCCGCTGCAAAATAGG 125
|||||
34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
|||||
126 TGGCCCTGGTAAAGGCTCCACCGACGGAGATCGGCTTCGCCATCGCCG 175
|||||
51 ArgLeuAlaGlnAspGlyAlaHisValValValSerSerArgGlyGlnI 67
|||||
176 GGTGGCCCGAGAGCGGCGCCATGTGTGTGTCAGAGCCGAGAGACAGA 225
|||||
67 naSValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
|||||
226 GAATGTGGACCAAGCGGTGGCCAGCTGCAGGGGAGAGGGCTAGCGCTGA 275
|||||
84 hrcGlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
|||||
276 CCGGACACCGTGTCCATGTGGAGAGCGGAGACCGGAGCGGTGTGGT 325
|||||
101 AlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerSnaI 117
|||||
326 GCGATGGCTGTGAAGCTTCATGGAGATATCGATATCTCATGTC 375
|||||
117 aAlaValAspProPhePheGlySerIleMetAspValThrGluGluValT 134
|||||
376 TGCTGTCAACCCCTTCTTGGAAGCCTAATGATGATCATGACGAGAGTGT 425
|||||
134 rPaSpLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
|||||
426 GGGACAAAGACTGTGACATTAATGTGAAGGCCACCCGATGACAAAG 475
|||||
151 AlaValValProGluMetGluLysArgGlyGlyGlySerValValIleVa 167
|||||
476 GCGAGTGTCCAGAAATGGAGAAACAGAGGCGGCTCATGTGTGTGCT 525
|||||
167 lSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnV 184
|||||
526 GTCTTCATATAGCAGCTTCAGTCCATCTCTGCTTCAGTCCCTTACAAAG 575
|||||
184 aSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
|||||
576 TCAGTAAACAGCCTTGCTGGGCTGACCAAGACCTGGCCATATAGAGCTG 625
|||||
201 AlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysTh 217
|||||
626 GCGCCCAAGGAACATTAGGTGAACCTCCCTAGCACCTGACTTATCAACAG 675
|||||
217 rSerPheSerArgMetLeuTrpMetAspLysGluLysGluLysMetL 234
|||||
676 TACCTTCACAGAGATCTCTGATGACAAAGAAAGAGAAAGATATGA 725
|||||
234 ySLuThrLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGly 250
|||||
726 AAGAAACCTCGCGATAGAGAGGTAGCGGAGGACAGAGATTTGTGTGGC 775
|||||
251 ILevalSerPheLeucysSerGluaspAlaSerTyrIleThrcGlyLurh 267
|||||
776 ATCGTCTCTTCTCTGCTCTGAAGATGCCAGTACATCATCGGGGAAC 825
|||||
267 rValValValGlyGlyThrProSerArgLeu 278
|||||
826 AGTGGTGGTGGTGGAGAACCCCGTCCGCCCTC 859
|||||

seq_name: gb_cm:AB062757
seq_documentation_block:
LOCUS AB062757 880 bp mRNA linear MAM 04-DEC-2001
DEFINITION Sus scrofa mRNA for carbonyl reductase/NADP-retinol
dehydrogenase.
complete cds.

```

ACCESSION AB062757
 VERSION AB062757.1 GI:17298118
 KEYWORDS
 SOURCE Sus scrofa male heart cDNA to mRNA.
 ORGANISM Sus scrofa
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 AUTHORS Hara, A., Imamura, Y., Abe, H., Inoue, S. and Ishikura, S.
 TITLE Molecular characterization of a carbonyl/retinal reductase of pig tissues
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 880)
 AUTHORS Hara, A.
 JOURNAL Direct Submission
 Submitted (06-JUN-2001) Akira Hara, Gifu Pharmaceutical University, Laboratory of Biochemistry, Mitahara-higashi 5-6-1, Gifu, Gifu 502-8585, Japan (E-mail: hara@ifu-pu.ac.jp, Tel: 81-58-237-8586, Fax: 81-58-237-8586).
 FEATURES
 source Location/Qualifiers
 1..880
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /sex="male"
 /tissue="heart"
 28..810
 /EC_number="1.1.1.184"
 /codon_start=1
 /product="carbonyl reductase/NADP-retinol dehydrogenase"
 /protein_id="BAB78528.1"
 /db_xref="GI:17298118"
 /translation="MASTGVERRRPLENKVALVYASTDGLIARIADODGAHYVS SRKQENDVTATIGEGLSYCTGVCHVGKAREPRLVAMAVNLHGVDILVSNALVN PPGNTIIDTEYWDKILHNNKATVLMTRAVVPEKKRGSGVLLVSVGADHPFN LQYINYSKTLGLTGNLVAELRNIRVNCILAPGLIKTNFSGVLMMDKARKREYMRES LRRLRGNEPDCAGIVSFICSEDAISYITGETVYVGGGTASRL"
 BASE COUNT 205 a 224 c 288 g 163 t
 ORIGIN
 alignment_scores:
 Quality: 1154.00 Length: 269
 Ratio: 4.490 Gaps: 0
 Percent Similarity: 95.539 Percent Identity: 83.271
 alignment_block:
 US-09-866-034-2 x AB062757 ..
 Align seg 1/1 to: AB062757 from: 1 to: 880
 10 CysAlaIaRgAlaTrpAsnSerValArgMetAlaSerSerGlyMetThrAr 26
 1 TGTTCACAGACGTGAGAGTGGTGGGACGACGACGCGGGGAGACCG 50
 26 gArGAspProLeuAlaAsnLysValAlaLeuValThrAlaSerThrAspG 43
 51 CCGAAACCGCTCGAGATAGGTGGCAGCTAGTAACAGCCCTCCACTGACG 100
 43 LylleGlyPheAlaIleAlaIaRgAlaRgLeuAlaGlnAspGlyAlaHisVal 59
 101 GGATCGGCTTCCTCGCCATCGCCGCGCTGCGCCAGATGGGGCCACGCTG 150
 60 ValValSerSerArgLysGlnGlnAsnValAspLlnAlaValAlaThrIle 76
 151 GTGCTCAGCAGCCGGAAGCAGAGATGTGACCGGACAGTGGACACT 200
 76 uGlnGlyGluGlyLeuSerValThrGlyThrValCysHisValGlyLysA 93
 201 GCAAGGGAGAGCGCTGAGCGTGAAGGCGACCGTGTCCACAGTGGGAAGG 250
 93 IacGAspArgGluArgLeuValAlaThrAlaValLysLeuHisGlyCly 109
 251 CCGAGGACCGGAGCGACTGCTGCGCTGACACTTCACGCGGGGC 300

110 IleAspIleLeuValSerAsnAlaAlaValAsnProPheGlySerI 126
 301 GTGGACATCTGCTGCTCCATGACAGCCGTCACCATCTTGTGAAACAT 350
 126 eMetAspValThrGluGluValThrAspLysThrLeuAspIleAsnVal 143
 351 CATGATGCTGAGGAGAGGTGGACAAAGATTTCTGACGTTAAACGCA 400
 143 ySalAlProAlaLeuMetThrLysAlaValAlaProGluMetGlyLysArg 159
 401 AGCCACAGCTCTCTGATGACAAAGCGAGTGTCCCGAGATGGAGAACCA 450
 160 GlyGlyGlySerValIleValSerSerIleAlaAlaPheSerProse 176
 451 GGAGGCGGCTCAGTGTGATCTATCCCTCCGTTGGAGCCTACCAACCAT 500
 176 rProGlyPheSerProTyrAsnValSerLysThrAlaLeuGlyLeuT 193
 501 TCCCATCTGGGCTTCAATGTTAGTAAACACCTTGTGCGGCTCA 550
 193 hTlyThrLeuAlaIleGluLeuAlaProArgAsnIleArgValAsnGly 209
 551 CCANAGACCTGGCCCTAGAGCTGGCCCAAGAACATTAAGGTGAACCTGC 600
 210 LeuAlaProGlyLeuIleLysThrSerPheSerArgMetLeuTrpMetAs 226
 601 CTGGCGCTGAGCTCATCAGACTTAACCTGACAGCGTGTGTGATGGA 650
 226 pLysGluLysGluGluSerMetLysGluThrLeuArgIleArgGluLeuG 243
 651 TAAGCAGAAAGAGGATGATGAAGAAAGAAAGCTGCAATCAAGAAAGTTAG 700
 243 lYgluProGluAspCysAlaGlyIleValSerPheLeuGlySerGluAsp 259
 701 GCAATCCAGAGAGACTGCGGCGCATCGTCTTCTTCTTCTTCTTCTTAAAT 750
 260 AlaSerTyrIleThrGlyGluThrValValAlaGlyGlyThrProse 276
 751 GCCACTTACATCAGTGGGAGACAGTGTAGTGGTGGAGGAGCGCATC 800
 276 rArgLeu 278
 801 CCGCCTC 807
 seq_name: gb_ro:AB045132
 seq_documentation_block:
 LOCUS AB045132 1312 bp mRNA linear ROD 05-DEC-2000
 DEFINITION Mus musculus mounRDR mRNA for NADPH-dependent retinol dehydrogenase/reductase, complete cds.
 ACCESSION AB045132
 VERSION AB045132.1 GI:11559413
 KEYWORDS
 SOURCE Mus musculus (strain:C57 black) liver cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Furukawa, A., Ohnishi, T., Huang, D., Araki, N. and Ichikawa, Y.
 TITLE cDNA cloning and characterization of peroxisomal short-chain dehydrogenase / reductase that reduce all-trans retinal to retinol
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1312)
 AUTHORS Furukawa, A., Huang, D., Ohnishi, T. and Ichikawa, Y.
 JOURNAL Direct Submission
 Submitted (23-JUN-2000) Aizo Furukawa, Kagawa Medical University, Faculty of Medicine, Department of Biochemistry, Miki-cho, Ikenobe 1750-1, Kita-gun, Kagawa 761-0793, Japan
 (E-mail: aifurukawa@kms.ac.jp; Tel: 81-87-891-2104, Fax: 81-87-891-2105)
 FEATURES
 source Location/Qualifiers
 1..1312
 /organism="Mus musculus"

```

/strain="C57 black"
/db_xref="taxon:10090"
/tissue_type="Liver"
115..897
/gene="mounRDR"
115..897
/gene="mounRDR"
115..897
/gene="mounRDR"
/codon_start=1
/product="NADPH-dependent retinol dehydrogenase/reductase"
/db_xref="GI:11559416"
/translation="MASSGCTRRNPISNKVALYATASDGIQFALARRLEAGAHVYS
SRKOQNDRAVATIQEGELSYTGIVCHVGRAREDEKLTTLAKRHOGIDILVNAAN
PFGNLMQVTEVWDKVLSTINVTATAMIKAVEMERKGGSVYIVGSAVGFTRPS
LGPYNVSKTALVGLTKNFAELAPKNIKRVNCLAIKTRFSSVLMEEKAREDEIKKA
MOIRLRKPEDCAGIVSFLCEDASYINGETVVVGGGPPSRL"
polya_site
1303..1312
BASE COUNT 328 a 305 c 385 g 294 t
ORIGIN

alignment_scores:
Quality: 1130.50 Length: 279
Ratio: 4.331 Gaps: 1
Percent Similarity: 93.548 Percent Identity: 80.287

alignment_block:
US-09-866-034-2 x AB045132

Align seg 1/1 to: AB045132 from: 1 to: 1312

1 MethiSLysAlaGly...LeuLeuGlyLeuCySAlaArgAlaTrpAsnSe 16
|||||
58 ATCGAGAAAGCGGACGGCTGTAGGTGTGACTCAACGCGTGCATGTC 107
|||||
16 rValArgMetAlaSerSerglyMethrArgArgAspProLeuAlaAsnL 33
|||||
108 GGTGGAGATGGCCAGTCCGGGTGACTGCGAAACCTCTCTCCAAATA 157
|||||
33 yValAlaLeuValThraLaserThraSPGLIleGlyPheAlaIleAla 49
|||||
158 AGGTGGCCCTGTGCACAGCTCCACCGAGGATGGCTTGGCCATCGCC 207
|||||
50 ArgArgLeuAlaGlnAspGlyAlaHisValAlaValSerSeraArgLysL 66
|||||
208 CGGCGCGTGGTGAAGATGGGGCCACGTCGTCTCAGCAGCCGCAACA 257
|||||
66 nGlnAsnValaSPGlnAlaValaThraLengInGlyGluGlyLeuSerV 83
|||||
258 CGAGATGTGACCGTGCACACTACAGGAGAGGCGCTGAGTC 307
|||||
83 aLThrGlyThraLcySHisValaGlyLysAlaGluAspArgGluArgLeu 99
|||||
308 TCACGTGGCATGCTGTGCCACGTGGGGAAGGAGAGATGAGAAAAGCTG 357
|||||
100 ValAlaThraAlaValLysLeuHisGlyLysIleAspIleLeuValSerAs 116
|||||
358 ATAACACGCGCTCTGACAGCTCACAGGGGATGTGATCTGTCTCCAA 407
|||||
116 nAlaAlaValaSPnProPheGlySerIleMetAspValThrGluGluV 133
|||||
408 TGCTCTGTCAACCTTCTTGGAAATCTAATGATGTGCACAGAGAGAG 457
|||||
133 aLTrpAspLysThraLysAspIleAsnValLysAlaProAlaLeuMetThr 149
|||||
458 TGTGGGACAAAGGTTTGAACATTAATGTGACAGCTACACCCATGATGAT 507
|||||
150 LysAlaValaValProGluMetGluLysArgGlyGlyGlySerValValAla 166
|||||
508 AAAGCGGTGTGCTCCAGATGTGAAAAGCAGAGAGGCGCTCAGTGTGAT 557
|||||
166 eValSerSerIleAlaIlePheSerProSerProGlyPheSerProTYA 183
|||||

```

```

558 TGTGGTTCGTAGCAGGCTTCACGTGTCCTCTCTGTCCTTACA 607
183 snValSerLysThraAlaLeuLeuGlyLeuThraLysThraLeuAlaIleGlu 199
|||||
608 ATTTAGCAAAACAGCTTTGCTGTCTTACTAAGAACCTTTCAGCGGAG 657
|||||
200 LeuAlaProArgAsnIleArgValaAsnCySLeuAlaProGlyLeuIleLys 216
|||||
658 TTGGCCCGGAGAAACATTCAGAGTCACTGTTACACCTGACTCATTCAA 707
|||||
216 sThrSerPheSerArgMetLeuTrpMetAspLysGlyLysGluSerM 233
|||||
708 GACTCATTCACGACATCTGTTGTGGCAGGAGAAACAGACAGACTTCA 757
|||||
233 eLysGluThraLysArgGlyLeuArgArgLeuGlyLysProGluAspCySAla 249
|||||
758 TAAAGAGCGCATGCAAAATCAGAGGCTAGGACGACGAGATGTGTCT 807
|||||
250 GlyIleValSerPheLeuCySergLysAspAlaSerTyrIleThrGlyL 266
|||||
808 GCGATAGTTTCCTTCTTATGCTCTGAAAGCCGACGTTACTCATGAGAG 857
|||||
266 uThrValValaValGlyGlyGlyThraProSerArgLeu 278
|||||
858 GACCGTAGTAGTGGGGGAGAGAACCCCTTCTCGCCTC 894
|||||

seq_name: gb_com:AB045133

seq_documentation_block:
LOCUS AB045133 1212 bp mRNA linear MAM 05-DEC-2000
DEFINITION Oryctolagus cuniculus rabnRDR mRNA for NADPH-dependent retinol
dehydrogenase/reductase, complete cds.
ACCESSION AB045133
VERSION AB045133.1 GI:11559415
KEYWORDS
SOURCE
ORGANISM Oryctolagus cuniculus liver cDNA to mRNA.
Oryctolagus cuniculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE
1 (sites)
Furukawa,A., Ohnishi,T., Huang,D., Araki,N. and Ichikawa,Y.
cDNA cloning and characterization of peroxisomal short-chain
dehydrogenase / reductase that reduce all-trans retinal to retinol
Unpublished
2 (bases 1 to 1212)
Furukawa,A., Huang,D., Ohnishi,T. and Ichikawa,Y.
Direct Submission
Submitted (23-JUN-2000) Atzo Furukawa, Kagawa Medical University,
Faculty of Medicine, Department of Biochemistry; Miki-cho, Ikenobe
1750-1, Kita-gun, Kagawa 761-0793, Japan
(E-mail:afurukawa@kms.ac.jp, Tel:81-87-891-2104, Fax:81-87-891-2105)
FEATURES
source
1..1212
/organism="Oryctolagus cuniculus"
/db_xref="taxon:9986"
/tissue_type="Liver"
1..783
/gene="rabinRDR"
1..783
/gene="rabinRDR"
/codon_start=1
/product="NADPH-dependent retinol dehydrogenase/reductase"
/protein_id="BAB1877.1"
/db_xref="GI:11559416"
/translation="MASSGVTRRDPLANKVAIVTASDGIQFALARRLEAGAHVYS
SRKOQNDRAVATIQEGELSYTGIVCHVGRAREDEKLTTLAKRHOGIDILVNAAN
PFGNLMQVTEVWDKVLSTINVTATAMIKAVEMERKGGSVYIVGSAVGFTRPS
LGPYNVSKTALVGLTKNFAELAPKNIKRVNCLAIKTRFSSVLMEEKAREDEIKKA
MOIRLRKPEDCAGIVSFLCEDASYINGETVVVGGGPPSRL"
polya_site
1203..1212
BASE COUNT 268 a 336 c 375 g 233 t
ORIGIN

```

	Quality: 1123.00	Length: 260
	Ratio: 4.492	Gaps: 0
	Percent Similarity: 96.154	Percent Identity: 85.385
inment_block:		
09-866-034-2 x AB045133		
gn seg 1/1	to: AB045133	from: 1 to: 1212
19	MetAlaSerSeGlyMetThrArgAspProLeuAlaAlaSerValAla	35
1	ATGGCCAGCTCGGGGTGACCCGCCGGACCCGCTCGCATTAAGGTGGC	50
35	aleuValThrAlaSerThrAspGlyIleGlyPheAlaAlaIleArgProL	52
51	CATGTAAGCGGCTCCACGACGGAGATCGGCTTGGCCATCGCCGGGTC	100
52	euaIaGlnAspGlyAlaHisValValaValSerSerArgLysGlnIAsn	68
101	TGGCCACAGATGGGGCCCGCATGTGCATACGACCGCGAAGACAGAAC	150
69	ValAspGlnAlaValAlaIleThrLeuGlnGlyGlyLeuSerValThrGl	85
151	GTGGACCGGGCGGTGGCTGCTACTGCGACGGCGGAGGGCTAGTGTACGG	200
85	yThrValCysHisValGlyLysAlaGlnAspArgGluArgLeuValAlaT	102
201	CACCGTGTGCCACGTGGGGAAAGCCGACGCGGGAACGGCTGTGCCA	250
102	hrlAlaValLysLeuHisGlyIleAspIleLeuValSerAsnAlaAla	118
251	CGGCCCTGAACCTCCACGAGGACATTGATTCCTGTCTCAATGCTGCA	300
119	ValAsnProPhePheGlySerIleMetAspValThrGluValArgAsp	135
301	GTCACCCCTTCTTGGAAAGCTAATGATGTCTGCACCGAGAGCTGTGGA	350
135	plyThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLysAlaV	152
351	CAGAGTTCTGGACATTAAACGTAAAGCCATGGCCCTGATGACAAAGCGG	400
152	alValProGluMetGluLysArgGlyGlySerValValIleValSer	168
401	TGTGTCCAGAGATGAGAAACGAGGAGCGGCTCCGTGTGATCGTGGCC	450
169	SerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnValSe	185
451	TCCATTAGACGCTTCATTCATTACGTGCGCCGTGGCCCTTATATGTATG	500
185	rLyThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeuAlaP	202
501	TAAACACGCTTGGTGGGCTTACCAAGAACCTGGCCCTGGAGGTGGCTG	550
002	roArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysThrSer	218
551	CCCGAAGACATCCGGGTGAATGCGCTGGGCACTCATCAAGACACAC	600
219	PheSerArgMetLeuTyrPheCAspLysGluLysGluSerMetLysGl	235
601	TTTCAGTAAAGCGGTGTGTGGAGACAAAGCACAAAGAGAAACATCATAC	650
235	uThrLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGlyIleV	252
651	AAAGTTGAGGATTAAGAAAGTTAAGGCAAAACCAAGAAATGTGGGGCATCG	700
252	alSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluThrVal	268
701	TGTCTTTCCTGTGCTCCCAAGACGCCACTACATCAGTGGGAGACAGTG	750
269	ValValGlyGlyGlyThrProSerArgLeu	278

US-09-866-

US-09-866-

Align seg 1/1 to: BC003484 from: 1 to: 1234

```

7 LeuGlyLeuCysAlaArgAlaTrpAsnSerValArgMetAlaSerSerg 23
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
2 TTAGGTGGCTGGACCTCAAGCGTGATGTGGTGAGATGGCCGCTCCG 51
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
23 yMetThrArgArgAspProLeuAlaAsnLysValAlaLeuValThrAla 40
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
52 GTTGAATCTCTCGAAACCTCTCTCGAATAAGGTGGCTGTGTCACAGCCT 101
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
40 eThraspGlyIleGlyPheAlaIleAlaArgArgLeuAlaGlnAspGly 56
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
102 CCACGACGCGGATGGCTTGGCATGCCCCGCTGCTGCTGAGAGATGG 151
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
57 AlaHisValValSerSerArgLysGlnGlnAsnValAspGlnAlaVal 73
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
152 GCCACAGCTGTGTCAGCAGCGCAACAGCAGCAAGTGTGACCGTCACT 201
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
73 lAlaThrLeuGlnGlyGlyLeuSerValThrGlyThrValCysHisVal 90
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
202 GCCACACTACAGAGGAGCGCTGAGTGTGATGTCATGTCATGTCGACAG 251
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
90 aGlyLysAlaGluAspArgGluArgLeuValAlaThrAlaValLysLeu 106
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
252 TGGGGAGGCAAGAGGATCGAGAAAGCTGATACCAACGCGCTGAGAGCGT 301
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
107 HisGlyGlyIleAspIleLeuValSerAsnAlaAlaValAspProPheP 123
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
302 CACCGGGGATGATGATCTCTGCTCCATGCTGCTGATACCCCTTTCTT 351
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
123 eGlySerIleMetAspValThrGluGluValTrpAspLysThrLeuAsp 140
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
352 TGGAAATCTAATGATGTGTACAGAGAGGTGTGGGCAAGGTTTGGACCA 401
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
140 lAsnValValAspAlaProAlaLeuMetThrLysAlaValAlaProGluMet 156
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
402 TTAATGTGACACTACAGCATGATGATTAAGCGGTGTGTCAGAGATG 451
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
157 GluLysArgGlyGlySerValValIleValSerSerIleAlaAlaPhe 173
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
452 GAAAGCGAGGAGCGGCTCAGTGTGATTTGGGTTGTCAGAGCGCTT 501
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
173 eSerProSerProGlyPheSerProTyrAsnValSerLysThrAlaLeu 190
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
502 CACTGCGTTCCTCTCTGGGTCTTACAAATGTAGCAAAACAGCTTTC 551
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
190 euGlyLeuThrLysThrLeuAlaIleGluLeuAlaProArgAsnLeuArg 206
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
552 TGGGTCTTACTAAGAACTTTGCAAGCGGAGTTGGCCCGAAGAACATTCGA 601
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
207 ValAsnCysLeuAlaProGlyLeuIleLysThrSerPheSerArgMetLe 223
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
602 GTGAAGCTGTAGCACCTGACTCATCAAGACTTCGATTCAGAGTGTGTT 651
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
223 uTrpMetAspLysGluLysGluLysSerMetLysGluThrLeuArgIleAla 240
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
652 GTGGAGAGGAAAGCAAGAGAGACTTCAATAAAGAGCATGCAAAATCA 701
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
240 rGArgLeuGlyLysProGluAspCysAlaGlyIleValSerPheLeuArg 256
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
702 GAAAGCTAGGCAAGCAGAGATTTGTGCTGGCATAGTTTCTTATATCC 751
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
752 TCTGAAGAGCGCAGTTACATCAATGAGAGAGACCGTAGTAGTGGGGAGAG 801
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
273 yThrProSerArgLeu 278
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
802 AACCCCTTCTGCGCTC 817
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: gb_ro:AB062758
seq_documentation_block:

```

```

LOCUS      AB062758                806 bp    mRNA    linear    ROD 04-DEC-2001
DEFINITION Rattus norvegicus mRNA for carbonyl reductase/NADP-retinol
ACCESSION  AB062758
VERSION    AB062758.1 GI:17298120
KEYWORDS   Rattus norvegicus (strain:Wistar) male heart cDNA to mRNA.
ORGANISM   Rattus norvegicus
REFERENCE  1 (sites)
AUTHORS    Hara,A., Imanura,Y., Abe,H., Inoue,S. and Ishikura,S.
TITLE      Molecular characterization of mammalian carbonyl/retnal reductases
JOURNAL    Unpublished
AUTHORS    2 (bases 1 to 806)
Hara,A.
FEATURES   Direct Submission
SOURCE     Submitted (06-JUN-2001) Akira Hara, Gifu Pharmaceutical University,
           Laboratory of Biochemistry, Mitahara-higashi 5-6-1, Gifu, Gifu
           502-8585, Japan (E-mail:hara@ifu-pu.ac.jp, Tel:81-58-237-8586,
           Fax:81-58-237-8586)
           Location/Qualifiers
             1..806
               /organism="Rattus norvegicus"
               /strain="Wistar"
               /db_xref="taxon:10116"
               /sex="male"
               /tissue_type="heart"
               /EC_number="1.1.1.184"
               /EC_start=1
               /product="carbonyl reductase/NADP-retinol dehydrogenase"
               /protein_id="BAB78529.1"
               /db_xref="GI:17298121"
               /translation="MASSGLTRONPLANKVALYASTDGLIARLRLEADGAHVVIS
               SRKQONVRAVALTQEGELSVTVGVCHVKADEDEKLIHNMALKHOGIDILVSAAVN
               PFGNLMQVTEEVNKKVLSINVTASAMIKAVPAMEKRGGSVTVISVAVPLRFS
               LGPNVSKTALIGLTKNPAELAKNIRVNCRLAAGLIKTRHSSVLMKEAREEMIKET
               MOIRRLKPEKCVSIVSFLESDSYINGETVVGGGTPSL"
BASE COUNT 202 a 187 c 240 g 177 t
ORIGIN
alignment_scores:
  Quality: 1073.00      Length: 260
  Ratio: 4.398          Gaps: 0
  Percent Similarity: 93.846      Percent Identity: 81.154
alignment_block:
US-09-866-034-2 x AB062758
Align seg 1/1 to: AB062758 from: 1 to: 806
19 MetAlaSerSergLysMetThrArgArgAspProLeuAlaAsnLysValAl 35
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
1 ATGGCGCAGTTCCGGGCTGACTGCGCCAGAACCACTGCGCAATAGGTGGC 50
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
35 aleuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArgArgL 52
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
51 CCTGTGCACAGCCTCCACGAGGAGTTGGCTTGCATCGCCGCGCGCTC 100
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
52 euAlaGlnAspGlyAlaHisValValValSerSerArgLysGlnGlnAsn 68
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
101 TGGCTGAGGAGCGGGCCCGTGTGTCATCAGCAGCGGCAAAAGCAGAAAT 150
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
69 ValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValThrGly 85
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
151 GTGGACCGTGCAGTGGCCACACTCCAGGAGAGGAGCACTGATGTTACTGG 200
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
85 yThrValCysHisValGlyLysAlaGluAspArgGluArgLeuValAlaAT 102
  ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
201 CGTGTGTCATGTGGGGAGCGAGAGATGAGAGAAAGCTGTGTAACA 250

```

```

102 hralaValLysLeuHISglyIleAspIleLeuValSerAsnAla 118
119 ValAspProPhePheGlySerIleMetAspValThrGluValTyrPAs 135
301 GTCACACCTTTCTTTGGAAATCTAATGACGTCACAGAGAGAGGTGGAA 350
135 pLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLysAla 152
351 CAAGGTTTGAGCATTAATGTGACACCTTCAGCAATGATTAAGCGG 400
152 aLValProGluMetGluLysArgGlyGlySerValIleValSer 168
401 TGGTGGCAGGATGAAAGAGAGAGGCGGCTCGGTGATGTGAGT 450
169 SerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnValSe 185
451 TCCGTAGCAGGCTTCCTTATTCCTCTGCGCCCTTACATGTAG 500
185 rLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeuAla 202
501 CAAGACAGCTTCTGCTGGTCTCAGTAAGACTTGGCCGAGAGTGGCC 550
202 rArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysThrSer 218
551 CAAGACAGCTTCTGCTGGTCTCAGTAAGACTTGGCCGAGAGTGGCC 600
219 PheSerArgMetLeuThrMetAspLysGlyLysGluGluSerMetLysG 235
601 TTCAGAGAGTGTCTTGGAAGGAGAAAGCAAGAGAAATGATAAAGA 650
235 ThrLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGlyIle 252
651 AACCATGCAATTTAGAGCTTAGGCAAGCCAGAGGATGTGCGTATAG 700
252 aLSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluThrVal 268
701 TTCCCTTCTTGCTCTGGAAGATGCGATGACATGATGCGAGACAGTA 750
269 ValValGlyGlyGlyThrProSerArgLeu 278
751 GTGGTTGGGGGAGAACCCCTTCTCGCTC 780

seq_name: gb_pr:AY071856

seq_documentation_block:
LOCUS AY071856 1003 bp mRNA linear PRI 20-JAN-2002
DEFINITION Homo sapiens NADP-dependent retinol dehydrogenase short isoform
ACCESSION AY071856
VERSION AY071856.1 GI:18252126
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1003)
AUTHORS Du,J., Liu,G.F., Wang,B., Sun,T., Chang,Y.M., Vennuste,G., Zhu,L.,
Chen,K.M., Wang,G.L., Xu,X.L. and Huang,D.Y.
TITLE Cloning and characterization of short NADP-dependent retinol
dehydrogenase/reductase that reduces all-trans retinal to retinol
Unpublished
2 (bases 1 to 1003)
JOURNAL Du,J., Liu,G.F., Wang,B., Sun,T., Chang,Y.M., Vennuste,G., Zhu,L.,
REFERENCE Chen,K.M., Wang,G.L., Xu,X.L. and Huang,D.Y.
AUTHORS Direct Submission
TITLE Submitted (26-DEC-2001) Cell Biology, China Medical University,
JOURNAL No.92 Beier Road,Heping District, Shenyang, Liaoning 110001, China
FEATURES
Source 1..1003
/organism="Homo sapiens"

```

```

CDS
<
/db_xref="taxon:9606"
6..584
/codon_start=1
/product="NADP-dependent retinol dehydrogenase short
isoform"
/protein_id="AAL61824.1"
/db_xref="GI:18252127"
/translation="MHKAGLGLCARANNSSRMASGSTRDPLANKVALVYASTDGI
GAIARKLAQDAHVAVSSRRQONVDAVALQEGLSVTVCVCHGKAEDRERLVAT
AVKHGIDILVSNAAVNPFGSINDVIEVWDKLMMDKEESKETLRIRRLGEPE
DCAGIVSEFCSDDASYITGEFVVGGPSPRL"
3'UTR
polyA_signal
BASE COUNT 957..962
ORIGIN 230 a 250 c 322 g 201 t

alignment_scores:
Quality: 923.00 Length: 278
Ratio: 4.807 Gaps: 1
Percent Similarity: 69.065 Percent Identity: 69.065

alignment_block:
US-09-866-034-2 x AY071856

Align seg 1/1 to: AY071856 from: 1 to: 1003

1 MethisLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTyrAsnSerVa 17
6 ATGCACAAAGCGGGGCTAGGCTCTGTGCCGGGCTTGGAATTCGCT 55
17 ArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLys 34
56 GCGGATGCCAGCTCCGCGATGACCCGCCGAGCCCTCGCAATTAAGG 105
34 aLLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
106 TGGCCCTGTAAGCGCTCCACCGAGGATCGCTTCCCATGCGCCGG 155
51 ArgLeuAlaGlnAspGlyAlaHisValValAlaSerSerArgLysGln 67
156 CGTTGGCCGAGGAGCGGCCCATGTGCTGTCAGCAGCGGAGAGCA 205
67 AsnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
206 GAATGTGACACAGCGGCTGCCACGCTGCAAGGGGAGGCGCTGAGT 255
84 hGlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuVal 100
256 CGGGCACCGTGTCCATGTGGGAAGCGGAGGAGCGGCTGGTG 305
101 AlaThrAlaValLysLeuHISglyIleAspIleLeuValSerAsnAl 117
306 GCCACGGCTGTGAAGCTTCATGAGGATGCAATCTCATGCTCCATGC 355
117 aLValAsnProPhePheGlySerIleMetAspValThrGluValT 134
356 TCGTGCACACCTTCTTGGAGCATATGATGTCACAGAGAGGTGT 405
134 rAspThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
406 GGGCAGAG..... 413
151 AlaValAlaProGluMetGluLysArgGlyGlySerValIleVal 167
413 ..... 413
167 iSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnV 184
413 ..... 413
184 aLSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200

```

```

413 ..... 413
201 AlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleIleYstH 217
413 ..... 413
217 rSerPheSerArgMetLeuTrpMetAspIlySGlySGlyGluSerMetL 234
414 ..... CTCTGCATGACCAAGAAAAGAGAAAGCATCA 447
234 ySGluThrLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGly 250
448 AAGAAACCTCGCGGATAGAGAGGTAGCGAGACGAGATTGCTGGC 497
251 IleValSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluTh 267
498 ATGGTCTCTTCTGCTGCTGAAGTGCACGCTACATCATCGGGGAAAC 547
267 rValValValGlyGlyGlyThrProSerArgLeu 278
548 AGTGGTGGGTGGGAGAACCCCGCCGCTC 581

seq_name: gb-Dr:BC007339

seq_documentation_block:
LOCUS BC007339 1226 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, short-chain alcohol dehydrogenase family member,
clone MGC:15246 IMAGE:3833714, mRNA, complete cds.
ACCESSION BC007339
VERSION BC007339.1 GI:13938401
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1226)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcgabs-remail.nih.gov
Tissue Procurement: AMCC/DC/DMP
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbio.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

FEATURES
Source
1..1226
/organism="Homo sapiens"
/db_xref="locusid:10202"
/db_xref="taxon:9606"
/clone="MGC:15246 IMAGE:3833714"
/tissue_type="Skin, melanotic melanoma."
/clone_id="N1H-MGC-20"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
49..825
/codon_start=1
/product="short-chain alcohol dehydrogenase family member"
/protein_id="AAH07339.1"

```

```

/d_b_xref="GI:13938402"
/translation="MSTGIDRKGVLANRVAVVTGSGIFAIRRLARDGAHYIS
SRKQNVNRAMAKIQEGELSVAGIYCHVKAKLEHROLYAKALEHGGVDPLYCSGVN
PLVSGTSLSEQIMDKILSVNVKISFALLSOLPWRNRKRAVILISITAYNPVAL
GVYNSKTAALLGTRTLAELAPKDIRNCVPIGLIKDESKVFHGNESLMKNREHH
QVORIGSEDDAGVIFLCSFPASVNEENIAVAVSTRL"
BASE COUNT 271 a 279 c 398 g 278 t
ORIGIN

alignment_scores:
Quality: 833.50 Length: 272
Ratio: 3.302 Gaps: 3
Percent Similarity: 87.500 Percent Identity: 61.029

alignment_block:
US-09-866-034-2 x BC007339

Align seg 1/1 to: BC007339 from: 1 to: 1226
2 HisLysAlaGlyLeuLeuGlyLeu.CysAlaArgAlaTrpAsnSerVala 18
|||||
3 CACGAGCAGCGCTGCTTCATCCCTGCTAGCTT.....TCTGTGA 46
18 rGMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysVal 34
47 GGATGAGCAGCACCGGGATAGACGAGAGAGCGCTGCTGCTAACCGGGTA 96
35 AlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArgAr 51
97 GCCGTGTCACGGGGGTCCACACAGTGGATGGCTTTGGCATGCCCGGAGC 146
51 GluAlaGlnAspGlyAlaHisValValValaSerSerArgLysGlnGlnA 68
147 TCTGGCCCGGAGCGGGCCACGCTGCTATCAGCAGCCGGAACGACGAGA 196
68 snValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValThr 84
197 ACCTGACCGGGGCCATGGCCAGCTGCAGGGGAGCGGCTGAGTGGCGG 246
85 GlyThrValCysHisValGlyLysAlaGluAspArgGluArgLeuValAl 101
247 GGCATTGTGTCCACCTGGGAGAGGCTGAGACCGGACACCTGGTGGC 296
101 athrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAsnAla 118
297 CAAGGCCCTGAGCACTGAGGGGGCTGACACTTCTGGTGTGACGCGAG 346
118 lAlaValAsnProPhePheGlySerIleMetAspValThrGluValThr 134
347 GGATCAACCTCTGTGAGGAGCAGCTGCTGGGAGCAGTGCAGATCTGG 396
135 AspLysThrLeuAsnIleAsnValLysAlaProAlaLeuMetThrLysAl 151
397 GACACAGATCTTAAGTGTGACCTGAGATCCCAAGCCCTGCTGAGCCA 446
151 aValValProGluMetGluLysArgGlyGlyLysSerValValIleValS 168
447 GTTGTGCGCTTACATGAGACAGG...AGGGTGTCTGTGATCCCTGGTCT 493
168 eSerIleAlaIleAsnSerProSerProGlyPheSerProTyrAsnVal 184
494 CTTCATTGACACTTAATATCCAGTAGTGGCCCTGGGTGTCAACATGTC 543
185 SerLysThrAlaLeuLeuGlyLeuThrLysThrIleAlaIleGluLeuAl 201
544 ACCAAGACAGCGCTGTGGGTCTCACTAGAACACATGGGATTCGAGCTGGC 593
201 aProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleIleYstHs 218
594 CCCAAGAGCATCGCGGTAACTGCTGCTCAGAGATTAATCAAAACCTG 643
218 ePheSerArgMetLeuTrpMetAspIlySGlySGlyGluSerMetL 234

```



```

644 ACTTCACCAAGTGTTCATGAGATGAGTCTCTGGAAGAACTCAAG 693
235 G1uthrleuarg1leargargleu1gly1progluaspysalagly1 251
694 GAACATCATGAGCTGACGAGAGATGGGAGCTGAGAGCTGTGCAGAAAT 743
251 evalSerPheleucyserser1uaspalasertyr1lethrGly1uthrv 268
744 CGTGCTCTCTGCTGCTCTCCACATGCGACCTACGTACAGGGGAGAAC 793
268 alVal1Val1Gly1 272
794 TTGGGTGTCAGGC 807

seq_name: gb_pr:HSU31875
seq_documentation_block: 1442 bp mRNA linear PRI 29-NOV-1995
LOCUS HSU31875
DEFINITION Human Hep27 protein mRNA, complete cds.
ACCESSION U31875
VERSION U31875.1 GI:1079565
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1442)
Gabrielli, F., Donadel, G., Bensl, G., Heguy, A. and Mell, M.
A nuclear protein, synthesized in growth-arrested human
hepatoblastoma cells, is a novel member of the short-chain alcohol
dehydrogenase family
Eur. J. Biochem. 232 (2), 473-477 (1995)
96035881
JOURNAL MEDLINE
REFERENCE
2 (bases 1 to 1442)
Gabrielli, F.
Direct Submission
Biochemistry, University of Pisa, Via Roma 55, Pisa 56126, Italy
FEATURES
source
1..1442
location/Qualifiers
1..1442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="hepatoblastoma Hep27"
434..1276
/note="similar to Streptomyces violaceoruber granatcin
polyketide synthase putative ketosyl, Swiss-Prot
Accession Number P16542"
/codon_start=1
/product="Hep27 protein"
/protein_id="AA82048.1"
/db_xref="GI:1079566"
/translation="MLSAVKARGYGWPHPCARLSVRMSSTGIDRKGVLANKVAAVTGS
TSIGFPAIARLARDGAHVIVSSRKQNVDRAMKLOGEGISVAGIYCHVKAEDREG
LVAKALEHCGGVDFLVCSAGVNPVLSGLTGSQIMDKILSVNKPSPALLSOLPPT
ENRGAVALVSSIAVNPVVALGVYNSKTLALDELTRILAEIAPKDIRVCVPGIT
KTDRSKVPHGNEISLMKNPKHEHQRIGESDCAIGVIFLCSPPASIVNGENIYVAGI
STR."
BASE COUNT 322 a 369 c 440 g 311 t
ORIGIN
alignment_scores:
Quality: 824.50 Length: 263
Ratio: 3.569 Gaps: 2
Percent Similarity: 87.833 Percent Identity: 61.597
alignment_block:
US-09-866-034-2 x HSU31875
Align seg 1/1 to: HSU31875 from: 1 to: 1442
10 CysAlaArgAlaTrpAsnSerValArgMetAlaSerSerGlyMetThrAr 26

```

```

479 TGTGTCTAGGCTT.....TCTGTAGGATGTAGACACCGGATAGACG 522
26 gArgAspProLeuAlaasnlyValAlaLeuValThrAlaSerThrAsp 43
523 GAAGCGGCTCTGCTAACCGGTAAGCGGTGTACGGGTCACACGATG 572
43 lylleGlyPheAla1leAlaArgArgLeuAlaGlnAspGlyAlaHisVal 59
573 GGATCGGCTTGTCCATGCCCGGAGCTGTGGCCGGGAGCGGCCACG 622
60 ValAlaSerSerArgLysGlnGlnAsnValAspLysAlaValAlaThrLe 76
623 GTCAATCAGACGCGGAACAGACAGATGGACCGGACGAGCGCAAGCT 672
76 uGlnGlyGlnGlyLeuSerValThrGlyThrValCysHisValGlyLys 93
673 GCAGGGGAGGGGCTGAGTGTGGCGGCAATGTGTGTCACAGTGGGAAAG 722
93 laGlnAspArgGluArgLeuValAlaThrAlaValLysLeuHisGlyLys 109
723 CTGAGGACCGGAGACAGCTGTGTGCCAAGGCCCTGGAGACACTGTGGGG 772
110 lLeAsp1leLeuValSerAsnAlaAlaValAsnProPheNeg1Ser1 126
773 GTCACTTCTCTGTGTACAGCGAGGGGTCAACCTGTGTAGAGACAC 822
126 eMeAspValThrGlnGlnValAlaTrpAspLysThrLeuAsp1leAsnVal 143
823 TCTGGGACACAGTACAGACATCTGGGACAGATCTTAAGTGTACAGTGA 872
143 ySalAspProAlaLeuMetThrLysAlaValAlaProGlnMetGlyLysArg 159
873 AGTCCCGACCGCTGCTGTGACGACGCTGCTGCCCTCATGTAGAGAACAG 922
160 GlyLysSerValVal1leValSerSer1leAlaAlaPheSerProse 176
923 ...AGGGGTCTGATCTGTGTCTCTTCATTCAGCTTAATATCCAGT 969
176 rProGlyPheSerProTyraAsnValSerLysThrAlaLeuLeuGlyLeu 193
970 AGTGGCGCTGGGTGTACAAATGTACAGACAGACAGCGCTCTGGGTCTCA 1019
193 hLysThrLeuAla1leGlnLeuAlaProArgAsn1leArgValAsnCys 209
1020 CTAGAACACTGGCATGTGAGCTGCCCCCAAGACATCCGGTAAACTGC 1069
210 LeuAlaProGlyLeu1leLysThrSerPheSerArgMetLeuTrpMetAs 226
1070 GTGTTCCAGGAATTATAAAACTGACTTCAGCAAGGTGTTTCATGGGA 1119
226 pLysGlyLysGlnGlnSerMetLysGluThrLeuArgGileArgArgLeu 243
1120 TGAGTCTCTCTGAGAGACTCTAAGAGACATCATCAGCTGTCAAGAGATTG 1169
243 Lys1uProGlnAspCysAlaGly1leValSerPheLeuCysSerGluAsp 259
1170 GGGAGTCAGAGAGACTGTGCAGGAATCGTGTCTCTCTGCTCTCCAGAT 1219
260 AlaSerTyrlleThrGlyGluThrValValValGly 272
1220 GCCAGCTACGTCAACGGGAGAACATTTGGGTGGCAGGC 1258

seq_name: gb_in:AY069779
seq_documentation_block: 1113 bp mRNA linear INV 17-DEC-2001
LOCUS AY069779
DEFINITION Drosophila melanogaster SD02021 full length cDNA.
ACCESSION AY069779
VERSION AY069779.1 GI:17862893
KEYWORDS FLI cDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

```


SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 34084)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
COMMENT This sequence was identified as CDM:1021081 by the submitter. For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source 1..34084
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 9185 a 7818 c 7622 g 9459 t
ORIGIN
alignment_scores:
Quality: 619.00 Length: 309
Ratio: 2.879 Gaps: 3
Percent Similarity: 69.579 Percent Identity: 44.984
alignment_block:
US-09-866-034-2 x AC015075/rev ..
Align seg 1/1 to reverse of: AC015075 from: 1 to: 34084
12 ArgAlaTTPasnservAlarGmetAlaSerSerglyMetThrArgArgas 28
17151 CCTTGTCCAGCAGTAGTCAGAAAGTTGCAGACGCCGACGATGAAGCGT... 17104
28 pProLeuAlaasnlyValAlaLeuValThrAlaSerThrAspGly... 43
17103 ...TTGGCAGAGAAAGTAGCCGTGTACCGCTCCACCGATGGGTAAAG 17058
43 43
17057 CAGCTCAATAGAACTGAGAGATGCCACCACTAATTCGATTATATC 17008
44 IleGlyPheAlaIleAlaArgArgLeuAlaGlnAspGlyAl 57
17007 CCTCCAGCATGTGGTTCCGATTTGCCAAAGCGTCGCCAGATGGCGC 16958
57 aHisValValAlaSerSerArgLysGlnGlnAsnValAspGlnAlaValA 74
16957 CCGATGTGTCATCAGCGATCGCAAGCAAAAGATGTGGACTCGCTCTGG 16908
74 laThrLeuGlnGlyLeuSerValThrGlyThrValCysHisVal 90
16907 CCGAGCTGGCGCAACGAACTGAAATGTCCACGCTCGAAGTCCATGTTC 16858
91 GlyLysAlaGlnAspArgGlnArgLeuValAlaThrAlaValLysLeuH 107
16857 ACCGAAACCGAGATCGCAACAGCTCTTGAAGAGACCATTAAGCAAGTT 16808
107 sGlyGlyIleAspIleLeuValSerAsnAlaAlaValAsnProPhePheG 124
16807 TGGCAAGCTCAACATTTGGTTCAGCAATGCAAGCAACCAATCTCGCGTGC 16758
124 lYSerIleMetAspValThrGlnGluValThrAspLysThrLeuAspIle 140
16757 GCGGTGTCTCCAGTCCGATGGAAGAGTGTGGACAGATCTTCGATGTG 16708
141 AsnValLysAlaProAlaLeuMetThrLysAlaValAlaProGluMetG 157
16707 AACGTGAAGAGTCTCTATCTGTCTGCGCAAGAGGACGACCTCTCTCTGCG 16658

157 uLysArgGlyGlyLyservAlaIleValSerSerIleAlaAlaPheS 174
16657 CCAGCAAAAGACTCCAGCATCGTTTCGTCTCTCATTTGCTGGCTATG 16608
174 eProSerProGlyPheSerPro..... 181
16607 ATGCTTTGA.GGTTACTGCTTATCTCGGTTCAAGTTGTCATGCTT 16559
182 TyrAsnValSerly 186
16558 ATCTCCGTATCTTATTCGATACCTACTGAGAGCTATTTCGTCACAA 16509
186 sThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeuAlaPro 203
16508 GACCGCGCTATTGGCTTGACCAAGCGACCGCCCAAGATCTGGCGCGG 16459
203 rGsnIleArgValAsnCysLeuAlaProGlyLeuIleLysThrSerPhe 219
16458 AGGCGATTCGCTCACTGCTGCTCCAGAGATCATCAGACAAAGTTTC 16409
220 SerArgMetLeuThrPheAlaAspLysGluLysGlnSerMetLysGlu 236
16408 TCGAAACACTGTACGAAATGAGTCGCGCAATGAGCGGCTTTGAGCAA 16359
236 rLeuArgIleArgArgLeuGlyLysProGluAspCysAlaGlyIleValS 253
16358 AATACCATGTGCTGTCTGGCACCAAGTGAAGATGCGTGGCGGTGCT 16309
253 ePheLeuGlySerGlyAspAlaSerTyrIleThrGlyLysThrValVal 269
16308 CTTTGTGTCTCCGAGGATGCGGCTTACATTACGAGATGCTCATTTGT 16259
270 ValGlyGlyGlyThrProSerArgLeu 278
16258 GCTGGGCTGATGATGACAGCTCTTTG 16232

alignment_block:
US-09-866-034-2 x AAZ42131

Align seg 1/1 to: AA42131 from: 1 to: 1302

```

1 MethHisLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVal 17
34 ATGCACAAAGCGGGGCTCTAGGCTCTGCTGCGGGCTTGGAATTCGGT 83
17 LArgMetAlaSerSerglyMetThrArgArgAspProLeuAlaAsnLysV 34
84 GCGGATGGCCAGCTCCGGGATGACCCCGGAGCCCGCTCGCAATAAAG 133
34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
134 TGGCCCTGGTAAAGCGCTCCACCGAGGAGATGCGCTTCGCCATCGCCGG 183
51 ArgLeuAlaGlnAspGlyAlaHisValValLysSerArgLysGlnG 67
184 CTTTGGCCAGGAGCGGCGCATGCTGCTGACAGCAGCGGAGAGAGA 233
67 nasnValAspGlnAlaValAlaThrLeuGlnGlyLeuSerValT 84
234 GAATGTGGACCAAGCGGCTGCGACGCTGCGAGGGAGGGCTGAGCGTGA 283
84 hGlyThrValCysHisValGlyLysAlaGluAspArgLysArgLeuVal 100
284 CGGCGACCGCTGCGCATGCTGGGGAAGCGGAGCGGAGCGGCTGCTG 333
101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerAsnAl 117
334 GCGAGCGGCTGAGAGCTTCATGAGATGATGATTCATGCTCCATGTC 383
117 aAlaValAsnProPheGlySerIleMetSerPalaIleArgLysValT 134
384 TGCTGTCAACCTTCTTGGAGCATATGATGATGATGATGAGGAGTGT 433
134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
434 GCGACAGAGCTCTGGACATTATGTAAGCGCCGACCCCTGATGACAAAG 483
151 AlaValAlaProGluMetGluLysArgGlyGlySerValIleIleVal 167
484 GCAGGCGGCTCCAGAAATGAGAAAGAGAGCGGCTCAGTGTATGCT 533
167 LserSerIleAlaIleAlaPheSerProSerProGlyPheSerProTyrAsnV 184
534 GCTTTCATAGCAGCGCTTACGTCATCTCTGCTTCACTCTTCAATG 583
184 aLserLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
584 TCAGTAATAACAGCTTGGTGGGCTGAGCAACAGACCTGGCCATAGAGCTG 633
201 AlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysTh 217
634 GCCCAAGAGACATTAGGCTGAGACCTGACAGCAGCTGAGCTTATCAAGAC 683
217 rSerPheSerArgMetLeuTrpMetAspLysLysGluGluGluSerMetL 234
684 TACGCTTCACAGAGATGCTTGCATGACAGCAAGAAAGGAGAGAGCATGA 733
234 ySGluThrLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGly 250
734 AAGAAACCTCGCGATTAAGAGAGTTAGCGGAGCAGAGAGATGTCTGCTGC 783
251 LLevalSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluTh 267
784 ATCGGTCTTCTTCTGCTGCTCAAGATGCCACATCATCATCGGGAAC 833
267 rValValIleGlyGlyGlyThrProSerArgLeu 278
834 ACTGTGTGTGGTGGAGAAACCCCGCTCCGCTC 867

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH14816
seq_documentation_block:

ID AAH14816 standard; cDNA, 1258 bp.
XX
AC AAH14816;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12619.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036;
PR 27-AUG-1999; 99JP-0300253;
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PR primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12619; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC of an oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SO Sequence 1258 bp; 282 A; 322 C; 394 G; 260 T; 0 other;

alignment_scores:

Quality: 1395.00 Length: 278
Ratio: 5.018 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.640

alignment_block:

US-09-866-034-2 x AAH14816

Align seg 1/1 to: AAH14816 from: 1 to: 1258

XX	AAZ46090:		
XX	05-MAY-2000	(first entry)	
XX			
XX			
DE	Short chain alcohol dehydrogenase-related molecule ScRM-1 cDNA.		
XX			
KW	Human: short chain alcohol dehydrogenase-related molecule;		
KW	SCAD-related molecule; ScRM-1; ScRM-2; metabolic regulator;		
KW	cell proliferation regulator; inflammation regulator;		
KW	cell proliferative disorder; immune disorder; arteriosclerosis;		
KW	atherosclerosis; bursitis; cirrhosis; hepatitis; AIDS;		
KW	Addison's disease; adult respiratory distress syndrome; allergy;		
KW	ankylosing spondylitis; amyloidosis; cancer; ss.		
XX			
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
FT	CDS	31..867	
FT		/*tag= a	
FT		/product=	"short chain alcohol dehydrogenase-related
FT			molecule ScRM-1"
XX	WO200004135-A2.		
PN			
XX			
XX	27-JAN-2000.		
PD			
XX			
PF	16-JUL-1999;	99MO-US16164.	
XX			
PR	16-JUL-1998;	980S-0116750.	
XX			
XX	16-JUL-1998;	980S-0160074.	
PA			
XX	(INCYTE) INCYTE PHARM INC.		
PI			
XX	Bandman O, Tang YT, Corley NC, Azimzai Y, Baughn MR;		
XX			
DR	WPI; 2000-171266/15.		
XX			
XX	P-PSDB; AAY68735.		
XX			
PT	New short chain alcohol dehydrogenase polypeptides useful for		
PT	diagnosis, treatment and prevention of cell proliferative disorders		
PT	such as atherosclerosis, cirrhosis and cancers of various tissues		
PS	Claim 7; Fig 1A-D; 78pp; English.		
XX			
CC	The present sequence encodes a human short chain alcohol dehydrogenase		
CC	(SCAD)-related molecule designated ScRM-1. The specification also		
CC	describes ScRM-2. SCM proteins are metabolic, cell proliferation and		
CC	inflammation regulators. The ScRM polynucleotides and polypeptides are		
CC	used for treating or preventing a cell proliferative or immune disorder		
CC	in humans. Cell proliferative disorders include arteriosclerosis,		
CC	atherosclerosis, bursitis, cirrhosis, and hepatitis. Immune disorders		
CC	include AIDS, Addison's disease, adult respiratory distress syndrome,		
CC	allergies, ankylosing spondylitis, and amyloidosis. The vectors,		
CC	agonists, antagonists, antibodies and complementary sequences are also		
CC	used for treating the above conditions. The polynucleotides and		
CC	polypeptides are also used for treating cancers of various tissues		
CC	such as adrenal gland, bladder, bone, bone marrow, and brain.		
XX			
XX			
SO	Sequence 1280 BP; 301 A; 326 C; 393 G; 260 T; 0 other;		
	Alignment scores:		
	<Quality: 1327.00	Length: 278	
	Ratio: 4.879	Gaps: 0	
	Percent Similarity: 97.842	Percent Identity: 96.403	
	Alignment block:		
	US-09-866-034-2 x AAZ46090		
	Align seg 1/1 to: AAZ46090 from: 1 to: 1280		

```

||||| ||| ||||||||||||||| ||| ::|||
31 ATCCACATGGCCAGGCTGCTAGGCTCTGTGCTGGGACGAGATCGGT 80
17 LARGMETLASERSERGLYMETHRARGARGSPROLEUALASNLVY 34
||||| ||||||| ||||||| ::|||
81 GCGGATGGCCAGCTCAGATGATGACCCCGGACCCGCTACCAATRAGG 130
34 ALALEUVALTHRALASERTRASPGLYLEGLYPHEALALEIARG 50
||||| ||||||| ||||||| ::|||
131 TGGCCCTGGTAACGGCTCCACCGACGGGATCGGCTTCGCATCGCCGG 180
51 ARGLEUALGLASPSGLYALAHISVALVALSERSERARGYSLNGI 67
||||| ||||||| ||||||| ::|||
181 CGTTTGCCCGACGACGGCCACACGTGTCTGACGACCGGAGACGACGA 230
67 NASNVALASPSGLNALVALATHREUINGLYGLYLEUSEVALT 84
||||| ||||||| ||||||| ::|||
231 GAATGTGACACAGCGGCTGCCACGCTGCAGSGAGGGGCTAGCGTGA 280
84 HNGLYTHRVALCYSHISVALGLYLYSVALGLUSPARGLUARGLEUVAL 100
||||| ||||||| ||||||| ::|||
281 CCGGACACCTGTGCCATGTGGGAMAGCGGACCGGACCGGCTGTG 330
101 ALATHRALVALLYSLEUTHISGLYLYLEASPLILEUVALSERASNA 117
||||| ||||||| ||||||| ::|||
331 GCCACGCGCTGTGAGCTCATGTGAGTATCTTACTTCCAAATGC 380
117 AALAVALAASPROPHEPHEGLYSERILEMETASPVALTNRGLUVALT 134
||||| ||||||| ||||||| ::|||
381 TCGTGTCAACCCCTTCTTGAAGCATATGATGATCACTGAGAGAGTGT 430
134 RPASPLYSTRILEUNSPILLEASNVALLYSALAPROALALEUETHRLYS 150
||||| ||||||| ||||||| ::|||
431 GGGACAGACTCTGGACATTATGTGAAGGCCACCGCTGATGACAAAG 480
151 AALAVVALPROGLMETGLUSARGLYGLYLYSERVALLEVALLEVA 167
||||| ||||||| ||||||| ::|||
481 GCAGTGGTCCAGAAATGGAGAAACAGAGAGGGGCTGAGTGTGATCGT 530
167 LERSEIRLEALALAPHERPROSERPROGLYPHERPROTYRASYV 184
||||| ||||||| ||||||| ::|||
531 GCTTTCATAGACAGCCTTCAGTCATCTCTGGCTTCACTCCTTACAATG 580
184 AISERTLYSTRALALEULEUGLYLEUTHRLYSSTRILEUALILEGLU 200
||||| ||||||| ||||||| ::|||
581 TCAGTAAACAGCCTTGTGGCTTCACAAATACCTGGCCATAGAGCTG 630
201 ALAPROARGASNLLEARGVALASNCYSLEUALAPROGLYLEUILEYTH 217
||||| ||||||| ||||||| ::|||
631 GCCCAAGAGACATTAAGGTGAACCTCTTAGCACCTGGACTTATCAAGAC 680
217 RSEPHESERARGMETLEUTRPMETASPLYSGULYSGLUGUSERMET 234
||||| ||||||| ||||||| ::|||
681 TAGCTTCACAGAGATGCTCTGATGACAGAGAAAGAGAAAGCATGA 730
234 YSGLUTHREUARGILEARGARGLEUGLYGLUPROGLUSPCYSALAGLY 250
||||| ||||||| ||||||| ::|||
731 AAGAAACCCCTGGGATTAAGAAAGTTAGCGAGACGAGATGTGTGTCG 780
251 ILEVALSERPHELEUCYSERGLUSPALASERTYRILETHGLYLU 267
||||| ||||||| ||||||| ::|||
781 ATCGTGTCTTCTTCTGTGCTCTGAAGATGCCACTACATCAGTGGGAAC 830
267 RVALVALVALGLYGLYLYTHRPROSERARGLEU 278
||||| ||||||| ||||||| ::|||
831 AGTGGTGGTGGTGGAGGAAACCCGCTCCGCTC 864

```

seq_name: /SIDS1/gcgdata/geneseq-emb1/NA2001A.DAT:AAH07482

seq_documentation_block:

ID AAH07482 standard; CDNA: 680 BP.

XX

AAH07482;

```

XX 26-JUN-2001 (first entry)
DT Human CDNA clone (5'-primer) SEQ ID NO:4317.
XX
DE Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
KW Homo sapiens.
XX
OS EP1074617-A2.
XX
PN 07-FEB-2001.
XX
PD 28-JUL-2000; 2000EP-0116126.
XX
PF 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1: SEQ ID 4317; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH1742 represent human cDNA sequences; AAH92446 to
CC AAH35893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 680 BP; 148 A; 180 C; 216 G; 130 T; 6 other:

```

alignment_scores:

Quality: 976.00 Length: 218
Ratio: 4.670 Gaps: 2
Percent Similarity: 95.872 Percent Identity: 94.954

alignment_block:

US-09-866-034-2 x AAH07482

Align seg 1/1 to: AAH07482 from: 1 to: 680

1 MethISLVALAGLYLEUENGLYLYSVALARGALATRPASNSERVA 17

|||||


```
21 ATGCACAAAGCGGGCTGTAGGCTCTGTGCCCGGCTTGGAAATCGCT 70
17 IArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaAsnLysV 34
   |||||
71 GCGGATGGCCAGCTCCGGGATGACCCGCGGACCCCTCCCAATATAAG 120
34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
   |||||
121 TGGCCCTGCAAGCGCTCCACGACGGGATCGGCTTCGCATCGCCCGG 170
51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnI 67
   |||||
171 CGTTTGGCCCGAGGAGCGGGCCATGTGTCTGACAGCCGGAAGCAGCA 220
67 nasValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerValT 84
   |||||
221 GAATGTGACACGCGGCTGGCCACGTCGACAGGGGAGGGCTGAGCGCTGA 270
84 hrGlyThrValLysHisValGlyLysAlaGlnAspArgLysArgLeuVal 100
   |||||
271 CGGCGACCGCTGTGCCATGTGGGGAAGCGGAGGACCGGAGCGGCTGTG 320
101 AlaThrAlaValLysLeuHisGlyGlyIleAspIleLeuValSerAsnAl 117
   |||||
321 GCCACGGCTGTGAAGCTTCATGGAGGTATGATCTCCTAGTCCCAATGC 370
117 aAlaValAspProPhePheGlySerIleMetAspValThrGluGluValT 134
   |||||
371 TGGCTGTAACCCCTTTCTTGGAGCATATGATGATGCTCAGTGAAGAGGTGT 420
134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
   |||||
421 GGCACAGACTCTGCACATTAATGTGAANCCCCAGCCCTGATACAAAG 470
151 AlaValAlaProGluMetGluLysArgGlyGlyLysValValIleVal 167
   |||||
471 GCACTGTGTCAGAAATGAGAAACGAAGAGCGGCTCAGTGGATGATCGT 520
167 lSerSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsn 183
   |||||
521 GTCTTCATATAAGAACCTTCAGTCATCTCTGCTTCAGTCTTACAAAT 570
184 ValSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLe 200
   |||||
571 GTCAGTAAACAGCCCTTGTCTGGCTGACCAANCCCTGCGCATANAGCT 619
200 uAlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLys 216
   |||||
620 GGGCCCAAGGACATTTANGTGAACCTGCTAACACCTGACATWATCAAA 669
seq_name: /sids1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.ABA06587
seq_documentation_block:
ID ABA06587 standard; cDNA; 1292 BP.
XX
XX ABA06587;
AC
XX
XX 10-JAN-2002 (first entry)
DT
XX
XX Human CDNA SEQ ID NO: 253.
DE
XX
XX Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200154474-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01349.
PR
```

```
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUL-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218920P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
```

PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237033P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239335P.
 PR 13-OCT-2000; 2000US-239337P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241221P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241786P.
 PR 20-OCT-2000; 2000US-241787P.
 PR 20-OCT-2000; 2000US-241808P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 20-OCT-2000; 2000US-241826P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 08-NOV-2000; 2000US-246474P.
 PR 08-NOV-2000; 2000US-246475P.
 PR 08-NOV-2000; 2000US-246476P.
 PR 08-NOV-2000; 2000US-246477P.
 PR 08-NOV-2000; 2000US-246478P.
 PR 08-NOV-2000; 2000US-246523P.
 PR 08-NOV-2000; 2000US-246524P.
 PR 08-NOV-2000; 2000US-246525P.
 PR 08-NOV-2000; 2000US-246526P.
 PR 08-NOV-2000; 2000US-246527P.
 PR 08-NOV-2000; 2000US-246528P.
 PR 08-NOV-2000; 2000US-246532P.
 PR 08-NOV-2000; 2000US-246609P.
 PR 08-NOV-2000; 2000US-246610P.
 PR 08-NOV-2000; 2000US-246611P.
 PR 08-NOV-2000; 2000US-246613P.
 PR 17-NOV-2000; 2000US-249207P.
 PR 17-NOV-2000; 2000US-249208P.
 PR 17-NOV-2000; 2000US-249209P.
 PR 17-NOV-2000; 2000US-249210P.
 PR 17-NOV-2000; 2000US-249211P.
 PR 17-NOV-2000; 2000US-249212P.
 PR 17-NOV-2000; 2000US-249213P.
 PR 17-NOV-2000; 2000US-249214P.
 PR 17-NOV-2000; 2000US-249215P.
 PR 17-NOV-2000; 2000US-249216P.
 PR 17-NOV-2000; 2000US-249217P.
 PR 17-NOV-2000; 2000US-249218P.
 PR 17-NOV-2000; 2000US-249244P.
 PR 17-NOV-2000; 2000US-249245P.
 PR 17-NOV-2000; 2000US-249264P.
 PR 17-NOV-2000; 2000US-249265P.
 PR 17-NOV-2000; 2000US-249297P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 17-NOV-2000; 2000US-249300P.
 PR 01-DEC-2000; 2000US-250160P.
 PR 01-DEC-2000; 2000US-250391P.
 PR 05-DEC-2000; 2000US-251030P.
 PR 05-DEC-2000; 2000US-251988P.
 PR 05-DEC-2000; 2000US-256719P.
 PR 06-DEC-2000; 2000US-251479P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251858P.
 PR 08-DEC-2000; 2000US-251859P.
 PR 08-DEC-2000; 2000US-251989P.
 PR 08-DEC-2000; 2000US-251990P.
 PR 11-DEC-2000; 2000US-254097P.
 PR 05-JAN-2001; 2001US-259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruden SM;
 XX
 DR WPI; 2001-476161/51.
 DR P-PSDB; ABB10365.
 XX
 PT Isolated nucleic acid molecule encoding an inflammation-associated

PT polypeptide is used in preventing, treating or ameliorating a medical
 condition
 XX
 PS Claim 1; SEQ ID NO: 253; 859pp + Sequence Listing; English.
 XX
 CC The present invention provides human cDNAs, proteins and related genomic
 CC DNAs. These can be used in the treatment of neural, immune system,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal and proliferative disorders and inflammation. The present sequence
 CC is a cDNA of the invention.
 XX
 SO Sequence 1292 BP; 282 A; 339 C; 423 G; 238 T; 10 other:

alignment_scores:
 Quality: 906.00 Length: 278
 Ratio: 4.768 Gaps: 1
 Percent Similarity: 68.345 Percent Identity: 67.986

alignment_block:
 US-09-866-034-2 x ABA06587

Align seg 1/1 to: ABA06587 from: 1 to: 1292

1 MethiLysAlaGlyLeuLeuGlyLeuCysAlaArgAlaTrpAsnSerVa 17
 263 ATGCACAGGCGSRGCTCTGAGCCCTGCGCCGGCTTGCAATTCGGT 312
 17 largMetIAserSerGlyMethTrpArgArgAspProLeuAlaAsnLysV 34
 313 GCGGATGCCAGCTCCGGATGACCCCGGACCCGCTGCAAAATAGG 362
 34 aAlaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
 363 TGCCCTGTGTAAGCGCTCCACGAGCGGATGCGCTTCCCATCCGCCG 412
 51 ArgLeuAlaGlnAspGlyAlaHisValValIserSerArgLysGlnG 67
 413 CATTGGCCCGAGGAGCGGCCCACTGTGCTGCACGACCGGAAGCAGCA 462
 67 nasValAspGlnAlaValAlaThrLeuGlnGlyIleGlyLeuSerVal 84
 463 GAATGTGACACAGCGGCGCCACGCTGCAGGGGAGGGGCGCTGAGCTGA 512
 84 hGlyThrValCysHisValGlyIleGlyAlaGlnAspArgGluArgLeuVal 100
 513 CGGGACCGGTGCCATGTGGGAAGCGGAGGACCGGAGCGCTGCTG 562
 101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValIserAsnAl 117
 563 GCCAYGCGTGTGAAGCTCATGAGGTATCGATATCTGATCTCCAAATG 612
 117 aAlaValAsnProPhePheGlySerIleMetAspValThrGluValI 134
 613 TGCTGTCAACCTTCTTGTGAAGCMTAATGATGATGACAGGAGAGCTGT 662
 134 rPAspLysThrLeuAspIleAsnValIleAlaProAlaLeuMetThrLys 150
 663 GGGACAG..... 670
 151 AlaValAlaProGluMetGluLysArgGlyGlySerValIleLeuVal 167
 670 670
 167 IserSerIleAlaIleAlaPheSerProIserProGlyPheSerProTyAsnV 184
 670 670
 184 alSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeu 200
 670 670
 201 AlaProArgAsnIleArgValAsnCysLeuAlaProGlyLeuIleLysTh 217

670 670
217 rSerPheSerArgMetLeuTrpMetAspIysGluGluSerMetL 234
671CTGTGGATGGACAAAGAAAAAGGAAAGCATGA 704
234 ysgIuTrIleuArgIleArgArgIleuGlyIuPProGluAspCysAlGly 250
705 AAGAACCCTCGGATAGAAAGTTAGCGGAGCAGAGATTGTCTGCG 754
251 lIleValSerPheLeuCysSerGluAspAlaSerTyrIleThrGlyIuTh 267
755 ATCGTCTCTTCTGTCTGTCTGAGATGCCAGTACATCACTGGGAAAC 804
267 rValValValGlyGlyGlyThrProSerArgIleu 278
805 AGTGTGTGTGTGTGAGAAACCCGCTCCGCTC 838

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA540887

seq_documentation_block:

ID AA540887 standard; cDNA: 1292 BP.

AC AA540887;

DT 17-DEC-2001 (first entry)

DE cDNA encoding novel human enzyme polypeptide #103.

Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
ligase; hyperproliferative disorder; immunodeficiency disorder;
autoimmune disorder; neurological disorder; metabolic disorder;
inflammatory disorder; cardiovascular disorder; reproductive disorder;
blood-related disorder; infectious disorder; gene therapy; cytostatic;
anti arthritic; nephrotropic; anticoagulant; ss.

XX Homo sapiens.

PN WQ20015301-AA2.

XX 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US01239.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220966.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-024610.
 PR 08-NOV-2000; 2000US-024611.
 PR 08-NOV-2000; 2000US-024613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2000US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM:
 XX WPI: 2001.465566/50.
 DR P-PSDB: AAU23017.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 PT
 PS Claim 4; SEQ ID NO 113; 1180pp; English.
 PS
 XX The present invention relates to the isolation of novel human enzyme
 CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AA540785-AA541684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 1292 BP: 282 A; 339 C; 423 G; 238 T; 10 other;

alignment_scores:

Quality: 906.00 Length: 278
 Ratio: 4.768 Gaps: 1
 Percent Similarity: 68.345 Percent Identity: 67.986

alignment block:

US-09-866-034-2 x AA540887 ..

Align seg 1/1 to: AA540887 from: 1 to: 1292

1 MethisLysAlaGlyLeuLeuGlyLeuCysAlaArgLalThrpsnSera 17
 |||||
 263 ATGCACAGGCSNCGCTGAGCCTGTGCGGGCTGGAAATTCGT 312
 17 LArgMetAlaSerSerGlyMetThrArgArgAspProLeuAlaSnLys 34
 |||||
 313 GCGGATGGCCAGCTCCGGATGACCCCGGAGCCCTGCCAATRAGG 362
 34 aLaLeuValThrAlaSerThrAspGlyIleGlyPheAlaIleAlaArg 50
 |||||
 363 TGGCCCTGCTAAGCGCTCCACGAGGATCGGCTTCGCCATCCGCCG 412
 51 ArgLeuAlaGlnAspGlyAlaHisValValSerSerArgLysGlnG 67
 |||||
 413 CGTTGGCCAGAGCGCGCCAGTGTGCTGCAGCAGCCGGAAGCAGCA 462
 67 naSnValAspGlnAlaValAlaThrLeuGlnGlyGlyLeuSerVal 84
 |||||
 463 GATGTGACAGCAGCGCTGGCCACGCTCGAGGGAGGGCTAGAGCTGA 512
 84 hArgLysValCysHisValGlyLysAlaGlnAspArgGlnArgLeuVal 100
 |||||
 513 CCGGACCGCTGTCCATGTGGAGAGCGGAGCAGCGGAGCGGTGCTG 562
 101 AlaThrAlaValLysLeuHisGlyIleAspIleLeuValSerSna 117
 |||||
 563 GCCATGCTGTGAGCTTCATGAGGTATCGATATCTGATCCCATAGC 612
 117 aAlaValAsnProPhePheGlySerIleMetAspValThrGlnGlnVal 134
 |||||
 613 TGCTGTCAACCTTCTTGGAGCMTATGATGATGATGATGATGATGAT 662
 134 rPAspLysThrLeuAspIleAsnValLysAlaProAlaLeuMetThrLys 150
 |||||
 663 GGGACAGG..... 670
 151 AlaValValProGlnMetGlyLysArgGlyGlySerValValIleVal 167
 670
 167 LserSerIleAlaAlaPheSerProSerProGlyPheSerProTyrAsnV 184
 670
 184 aSerLysThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGlnLeu 200
 670
 201 AlaProArgAsnIleArgValAsnGlyLeuAlaProGlyLeuIleLysTh 217
 670
 217 rSerPheSerArgMetLeuTrpMetAspLysGlyGlnGlnSerMetL 234
 |||||
 671CTTGTGATGACAGCAAGAAAGAGCAAGCATGA 704
 234 ysgLthrLeuArgIleArgArgLeuGlyGlnProGlnAspCysAlaGly 250
 |||||
 705 AAGAAACCTCGGATGATAGAGGATGAGCGAGCAGCATTTGCTGGC 754
 251 lIleValSerPheLeuGlySerGlnAspAlaSerTyrIleThrGlyLys 267
 |||||
 755 ATCGTGTCTTCTTCTGCTGATGATGATGATGATGATGATGATGATG 804

267 rValValValGlyGlyThrProSerArgLeu 278
|||||
805 AGTGGTGTGGTGCAGCAACCCGCTCCGCCCTC 838

seq_name: /STDJ1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA529730

seq_documentation_block:

ID AA529730 standard; cDNA; 1292 BP.

AA529730;

21-NOV-2001 (first entry)

Human endocrine polypeptide encoding cDNA SEQ ID No 230.

Endocrine protein; human; mouse; rabbit; goat; horse; food additive;
cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
antiheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
cerebroprotective; neotropic; antibacterial; vitruicide; fungicide; cancer;
ophthalmological; vulnetaary; gene therapy; autoimmune disease; neoplasm;
hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;
cerebrovascular disorder; nervous system disorder; bacterial infection;
fungal infection; viral infection; ocular disorder; endocrine disorder;
gastrointestinal disorder; renal disorder; respiratory disorder;
wound healing; skin aging; organ transplantation; food preservative;
tissue regeneration; anti-infertility.

Homo sapiens.

WO200155364-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US01308.

31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227189.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235498.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.


```

419  ACRGAATTGAATCCACGCTTCGAAGTGGCATGTCCAGCGAACCCGAGG 468
      ::::::::::: ||| ::::::::::: ::::::::::: |||
95  spargluargleuvalaValaThraIaValaLysleuHslgIyIleasP 111
      ::::::::::: ||| ::::::::::: |||
469  ATCGCAACAGCTTTGAGGAGCACATTAGCAAGTTTGGCAAGGTCAAC 518
      ::::::::::: ||| ::::::::::: |||
112  ILeuValSerAsnAlaIaValaAsnProphheglIySerIleMetas 128
      ::::::::::: ||| ::::::::::: |||
519  ATTTGTGTGACCATGACGCCAACATCTCCGGTTCGGCGGTGTCTCGA 568
      ::::::::::: ||| ::::::::::: |||
128  pValThrgIgluValIrrpasplysThrleuAspIleasnValLysAlaP 145
      ::::::::::: ||| ::::::::::: |||
569  GTGCGATGAGAGCTGTGGCACAAGATCTTGAATGTGAACGTGAAGATT 618
      ::::::::::: ||| ::::::::::: |||
145  roAlaIeumetThrLysAlaValaValaProgluMetGluLysArgIyGly 161
      ::::::::::: ||| ::::::::::: |||
619  CCTATCTGCTGGCCAAAGAGGACACTGCTCTCTCGCGCCACCAAGAAGAC 668
      ::::::::::: ||| ::::::::::: |||
162  gLyservalValIleValSerSerIleAlaAlaIaIaIaIaIaIaIaIa 178
      ::::::::::: ||| ::::::::::: |||
669  TCCAGCATCTTTTCGTCTCTCTCATTTGCTGGCTATATGCTTTGAGCT 718
      ::::::::::: ||| ::::::::::: |||
178  yPheSerProTyraAsnValSerLysThrAlaIleuLeuGlyLeuThrLysT 195
      ::::::::::: ||| ::::::::::: |||
719  ACTGGAGGACTATTCCGTGACGAAGACGGCGCTGATTTGGCTTGACCAAG 768
      ::::::::::: ||| ::::::::::: |||
195  hIreunAlaIleGluLeuAlaIaProIArgAsnIleArgValaAsnCysLeuAla 211
      ::::::::::: ||| ::::::::::: |||
769  CACCCGCGCAAGATCTGGCCGCCGAGGAGCATCTGCTCACTGCTGTGCT 818
      ::::::::::: ||| ::::::::::: |||
212  ProGlyLeuAlaLysThrSerPheSerArgMetLeuIrrpetaSplysGI 228
      ::::::::::: ||| ::::::::::: |||
819  CCAGGAGTACTACGACACAAGTCTCAAGACACTGTACGAGAAATGAGTC 868
      ::::::::::: ||| ::::::::::: |||
228  uLysGluGluSerMetLysGluThrLeuArgIleArgArgLeuGlyLup 245
      ::::::::::: ||| ::::::::::: |||
869  GCGGATGTGAAGCGCTTTCGCAAAATATCCCATCGAGTGGTGTGGCAACA 918
      ::::::::::: ||| ::::::::::: |||
245  rogluAspCysAlaGlyIleValSerPheLeuLysSerGluAspAlaSer 261
      ::::::::::: ||| ::::::::::: |||
919  GTGAGGAGATGCTGGCGTGGCTCTTTTGGCTTCCGAGGAAGCGCGCG 968
      ::::::::::: ||| ::::::::::: |||
262  TyTleThrGlyIuThrValaValaIaIaGlyIyGlyThrProSerArgle 278
      ::::::::::: ||| ::::::::::: |||
969  TACATTAACCGAGAGATTCATTGTGGCTGGCGGTGGATGACAGCTCGTTT 1018
      ::::::::::: ||| ::::::::::: |||
278  u 278
      |
1019  G 1019

```

seq_documentation_block:	
ID	ABL16756 standard; DNA; 3147 BP.
XX	
XX	ABL16756;
AC	
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster genomic polynucleotide seq ID NO 1741.
XX	
KW	Drosophila: developmental biology; cell signalling; insecticide
KW	pharmaceutical; gene; ds.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	


```

PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1: SEQ ID NO 1741; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3147 BP; 884 A; 785 C; 726 G; 752 T; 0 other;

alignment_scores:
    Quality: 619.00      Length: 309
    Ratio: 2.879        Gaps: 3
    Percent Similarity: 69.579    Percent Identity: 44.984

alignment_block:
US-09-866-034-2 x AB16756/rev ..

Align seg 1/1 to reverse of: AB16756 from: 1 to: 3147

12 ArgAlaTrpAsnSerValArgMetAlaSerSerGlyMetThrArgAs 28
1923 CGCTTGTCACGAGTAGTCAAACTGCACAGCCGCGCATGAAGCT.. 1876
28 pProLeuAlaAsnLysValAlaLeuValThrAlaSerThrAspGly... 43
1875 ...TTGGCAGGAAAGTAGCCGTGGTCACCGCTCCACGATGGGTAG 1830
43 .....
1829 CAGCTCATAGAACTGAGAGAGATCGCACCCCACTAATCTGATTAATC 1780
44 .....11leGlyPheAlaIleAlaArgArgLeuAlaGlnAspGlyAl 57
1779 CCTCCAGAGATTGTTTCGATTCGCAAAAGCTGGCCGCAAGATGGCC 1730
57 aHisValValAlaSerSerArgLysGlnGlnAsnValAspGlnAlaVal 74
1729 CGCAGGTGTCATCGACATGCAAGCAAAAGATGTGACTCTGCTCTG 1680
74 laThrLeuGlnGlyGlyLeuSerValThrGlyThrValLysSHSVal 90
1679 CGAGAGTGGCAACACGAACTTAATCTCACCGCTCTGAAGTGCATGTC 1630
91 GlyLysAlaGlnAspArgLysArgLeuValAlaThrAlaValLysLeu 107
1629 ACGCAACCCGAGATGCAAAACGCTCTTGAAGAGACACATTAGCAAGTT 1580
107 sGlyGlyLysPheLeuValSerAsnAlaAlaValAsnProPhePheG 124
1579 TGGCAAGCTCAACATCTGTGTGACGCAATGACCCCAATCTCGCGTGG 1530
124 LysSerIleMetAspValThrGluGluValThrAspLysThrLeuAsp 140

```

```

1529 GCGGTGCTCTCGAGTCGAGAGAGAGTGTGGGACAAAGATCTTCGATGTC 1480
141 AsnValLysAlaProAlaLeuMetThrLysAlaValAlaValProGluMeG 157
1479 AACGTGAAGAGTCTCTATCTGTGCGCAAGAGAGAGCTGCTCTGCG 1430
157 uLysArgGlyGlySerValValIleValSerSerIleAlaAlaPheS 174
1429 CCAGCAAAAGAACTCCACATCGCTTTCGCTCTCCCTCATCTGCGTATG 1380
174 erProSerProGlyPheSerPro..... 181
1379 ATGCCCTTGA. GGTTAGTCGCCCTTAATCTCGGATTCAGTTTCATGCTT 1331
182 .....TyrAsnValSerLys 186
1330 ATTCGCCGTACTTAATTCGCAATAGTACTGAGAGCTTAATTCGTCAGCA 1281
186 sThrAlaLeuLeuGlyLeuThrLysThrLeuAlaIleGluLeuAlaPro 203
1280 GACCGCGCTGATGTGCTTGACCAAGGACGCGCCAGAGATGCGCGCCG 1231
203 rGAsnIleArgValAsnLysLeuAlaProGlyLeuIleLysThrSerPhe 219
1230 AGGCAATTCGCGTCACATCGCTGCTCCAGAGATCATCAGCAAAAGTTC 1181
220 SerArgMetLeuTrpMetAspLysGluLysGluGluSerMetLysGluTh 236
1180 TCGAAAGCACTGTACGAGAAATGAGTCGGCAATGAGCGCTTGGACAA 1131
236 rLeuArgIleArgArgLeuGlyGluProGluAspCysAlaGlyLysValS 253
1130 AATACCATGTCGCTGCTGGCGACACAGAGAGATGTCGCTGCTGTCT 1081
253 erPheLeuCysSerGluAspAlaSerTyrIleThrGlyGluThrValVal 269
1080 CTTTTTGTCTCCGAGATGCGGGCTACATTACGAGAGATGCATGTGTG 1031
270 ValGlyGlyGlyThrProSerArgLeu 278
1030 GCTGGCGGTGAATGACAGACGCTTTG 1004

seq_name: /STDS1/gcgdata/geneseq/geneseg-emb1/NA2001B.DAT.AB107994
seq_documentation_block:
ID AB107994 standard: cDNA: 2699 BP.
XX
AC AB107994;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18464.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.

```


PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143452.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

[illegible]

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication meaning no sequences were

[illegible]